
ENCADREMENT

Directeur de thèse : Dr. Thierry Lefrançois, directeur de l'unité mixte de recherché CIRAD-INRA contrôle des maladies animales exotiques et émergentes, Montpellier, France.

Co-directeur de thèse : Prof. Olivier Gros, Professeur à l'université des Antilles et de la Guyane, unité mixte de recherche CNRS-MNHN-UPMC 7138 Systématique-Adaptation-Evolution, équipe « Biologie de la mangrove », Pointe-à-Pitre, Guadeloupe.

Encadrant : Dr. Nathalie Vachieri, responsable de site, unité mixte de recherche CIRAD-INRA contrôle des maladies animales exotiques et émergentes, Petit-Bourg, Guadeloupe.

AUTRES MEMBRES DU JURY

Président : Prof. Abel Hiol, professeur à l'institut universitaire et technologique, département génie biologique, industries alimentaires et biologiques, université des Antilles et de la Guyane, Saint-Claude, Guadeloupe.

Rapporteur: Prof. Philippe Brouqui, professeur à la faculté de médecine de Marseille, unité mixte de recherche CNRS-IRD maladies infectieuses tropicales et émergentes, Marseille, France.

Rapporteur: Dr. Pascal Lebourgeois, maître de conférence à l'université Paul Sabatier Toulouse III, laboratoire de microbiologie et génétique moléculaire CNRS, équipe évolution des génomes bactériens, Toulouse, France.

Examineur: Dr. Bernard Mari, directeur de recherche à l'institut de pharmacologie moléculaire et cellulaire CNRS, université Nice Sophia Antipolis, équipe physiologie génomique des eucaryotes, Sophia Antipolis, France.

REMERCIEMENTS

Un grand merci au Dr Nathalie Vachier pour m'avoir permis de réaliser cette thèse sous son encadrement. Merci pour toute l'aide que tu m'as apporté tout au long de ces trois années. Merci pour ta confiance, ta disponibilité et tous tes conseils. A tes côtés, j'ai appris beaucoup de choses tant sur le plan professionnel que le plan personnel. Ta gentillesse, ta patience, ainsi que tout ton soutien m'ont été très précieux tout au long de ma thèse. Merci pour tout.

Je tiens à remercier le Dr Thierry Lefrançois pour son encadrement et ses conseils tout au long de ma thèse. J'ai été ravi de réaliser cette thèse au sein de l'équipe sous ta direction, merci pour tous les conseils fournis tout au long de ma thèse.

Je remercie également le professeur Olivier Gros d'avoir accepté de co-encadrer ma thèse durant ces trois années.

Je remercie le professeur Philippe Brouqui et le Docteur Pascal Lebourgeois qui m'ont fait l'honneur et le plaisir d'accepter d'être les rapporteurs de mon travail de thèse.

Je tiens également à remercier le professeur Abel Hiol, d'avoir accepté de présider à mon jury de thèse et d'avoir

Merci au Docteur Damien Meyer, pour tous les différents échanges que nous avons eu durant ma thèse, mais aussi pour les corrections des articles et des réunions de synthèse sur les résultats. Merci de tout l'intérêt porté à mes travaux.

Je n'oublie pas le Dr Pascal Barbry de l'IPMC à Sophia Antipolis pour m'avoir accueilli au sein du laboratoire. Un grand merci au Dr Bernard Mari pour tout son aide tout au long de ma thèse et plus particulièrement lors de mon séjour à Sophia Antipolis. Merci pour tout tant sur le plan professionnel que sur le plan amical. Grâce à toi, mon séjour à Sophia Antipolis restera un très bon souvenir. Une attention toute particulière à Sandra Fourré pour toute l'aide apportée pour la réalisation des hybridations. Merci également à Kevin Lebrigand pour le design des puces à ADN et pour son aide lors de l'analyse des lames. Je tiens à remercier sincèrement toute l'équipe de l'IPMC pour l'aide précieuse qu'ils m'ont apportée, mais aussi pour leur accueil chaleureux et leur bonne humeur. J'ai passé trois belles semaines à vos côtés.

Un grand merci à toute l'équipe de l'UMR CMAEE, en particulier à Christian Sheikboudou, pour toute l'aide qu'il m'a apportée durant ces trois

années, merci de m'avoir formé à la culture cellulaire, j'ai passé de bons moments à tes cotés. A Valérie Pinarello, pour toute l'aide qu'elle m'a apporté pour mes manips. A Yolaine Vincent, pour toute l'aide administrative qu'elle m'a apporté et pour avoir organisé ma soutenance de thèse comme un chef. Un merci également à Rosalie Aprelon, Cathy Carasco-Lacombe, Mélanie Magnan, Claire Fritsh et Laura Brudey pour leur bonne humeur et les discussions enrichissantes au bureau.

Un grand merci aussi au Dr Loïc Emboulé, pour toute l'aide qu'il m'a apporté durant ma thèse, pour tout son transfert de savoir, et pour tous ses conseils. Ton amitié ainsi que ta bonne humeur m'ont été très bénéfique tout au long de ma thèse.

Je tiens également à remercier tous les stagiaires et VCAT que j'ai pu côtoyer tout au long de la thèse. Une pensée particulière pour Pryscélia Gely pour toute son aide sur mon sujet de thèse ainsi que pour la rédaction de mon 1^{er} article. Ta bonne humeur et ta joie de vivre m'ont beaucoup apporté durant ces 6 mois de stages. Je garderai un très bon souvenir de ton passage.

Enfin, je tiens à remercier toute ma famille pour tout le soutien qu'ils ont su m'accorder durant ces trois années. Merci pour vos encouragements qui m'ont permis de ne jamais baisser les bras.

Pour finir, un grand merci à Sabrina, qui depuis toutes ces années est à mes cotés. Tu m'as toujours soutenu et cru en moi. C'est sans hésitation que tu m'as suivi en Guadeloupe, je te dédie cette thèse.

RESUME

Au cours de la thèse, l'étude du transcriptome de souches Gardel et Senegal virulentes et atténuées d'*E. ruminantium* a été réalisée. Une analyse du transcriptome à différents stades de développement, a d'abord été effectuée pour la souche Gardel virulente. Au stade corps réticulé (forme intracellulaire non infectieuse), une surexpression des gènes codant pour des protéines impliquées dans le métabolisme, le transport et l'échange de nutriments et dans la résistance au stress oxydatif était observée. Il semblerait que *E. ruminantium* mette en place un panel de mécanismes pour sa survie et son développement à l'intérieur de la cellule hôte. Au stade corps élémentaire (forme extracellulaire infectieuse), le gène *dksA* codant pour un facteur de transcription était surexprimé. Ce gène a été montré comme étant impliqué dans la régulation de facteurs de virulence. Il semblerait donc, qu'au stade corps élémentaire, il y ait une induction de mécanismes de virulence.

La comparaison de l'expression des gènes au stade corps élémentaire entre souches virulentes et atténuées a aussi été effectuée. Nos résultats ont montré une modification importante de la membrane pour les souches virulentes et atténuées. Pour les souches atténuées, il a été montré une surexpression des gènes impliqués dans la biogénèse membranaire et une sous-expression des protéines de la famille multigénique *map*. Ces résultats suggèrent que les protéines MAP jouent un rôle de leurre vis-à-vis de la réponse immunitaire protectrice. Des protéines membranaires hypothétiques sont surexprimées à la fois chez les souches virulentes et atténuées. Certaines d'entre elles surexprimées chez les souches atténuées semblent être de bons candidats vaccinaux et devraient être étudiées.

Une surexpression des gènes impliqués dans le métabolisme a été montrée pour les souches atténuées suggérant une adaptation spécifique des bactéries à leur cellule hôte. Nous avons aussi établi un lien entre les

modifications d'expression de certains gènes et des mutations sévères affectant ces gènes. Un gène *recO* surexprimé chez Senegal virulente est tronqué chez l'atténuée. Ce gène est connu pour être impliqué dans la réparation de l'ADN et lorsqu'il est muté induit l'atténuation chez *Rickettsia prowazekii*. Il pourrait être responsable de l'accumulation de mutations observées chez Senegal atténuée.

ABSTRACT

Transcriptomic study of Gardel and Senegal both virulent and attenuated *E. ruminantium* strains was conducted during my PhD. An analysis of transcriptome at different stages of development has been first conducted for virulent Gardel strain. At reticulate body stage (intracellular form non-infectious), over-expression of genes coding for proteins involved in metabolism, transport and exchange of nutrients and resistance to oxidative stress was observed. At this stage of development, *E. ruminantium* seems to activate mechanisms for its survival and development within the host cell. At elementary body stage, *dksA* the gene encoding for a transcription factor was over-expressed. This gene has been shown to be involved in the regulation of virulence factors. It seems, therefore, at the elementary body stage, *E. ruminantium* induces its virulence factors.

Secondly, we compare the transcriptome of elementary body between virulent and attenuated strains. Our results showed an important membrane modification of attenuated and virulent strains. For attenuated strains, we observed an over-expression of genes involved in membrane biogenesis and a diminution of expression of *map* multigenic family. It seems that MAP proteins subvert the protective immune response. Hypothetical membrane proteins are over-expressed in both virulent and attenuated strains. Some over-expressed proteins in attenuated strains, seem to be good vaccine candidates and will be studied.

Metabolism genes were over-expressed in both attenuated strains, suggesting specific adaptation of bacteria to their host cells. There is a link between gene expression modifications and severe mutation affecting these genes. *RecO* gene over-expressed in virulent Senegal was truncated in attenuated one. It is involved in DNA repair and mutated *recO* induces attenuation of *Rickettsia prowazekii*. It could be responsible of the accumulation of mutations in attenuated Senegal.

SOMMAIRE

Chapitre 1 - Introduction	1
<u><i>Partie 1</i></u>	
La cowdriose	2
<u><i>Partie 2</i></u>	
Agent causal de la cowdriose : <i>Ehrlichia ruminantium</i>	12
<u><i>Partie 3</i></u>	
Caractérisations des interactions hôtes-pathogènes-vecteurs par des approches intégratives à haut débit	18
<u><i>Partie 4</i></u>	
Objectifs de la thèse	85
Chapitre 2 - Stratégie d'étude et protocoles expérimentaux mis en œuvre pour l'analyse du transcriptome d'<i>Ehrlichia ruminantium</i>	
<u><i>Partie 1</i></u>	
Production des réplicats biologiques Gardel et Sénégal	88
<u><i>Partie 2</i></u>	
Méthode SCOTS	91
<u><i>Partie 3</i></u>	
Analyse du transcriptome d' <i>Ehrlichia ruminantium</i> par microarrays	94
Chapitre 3 - Résultats	
<u><i>Partie 1</i></u>	
Réplicats biologiques Gardel et Senegal produits	102

<u>Partie 2</u>	
Analyse du transcriptome des souches Gardel et Senegal virulente et atténuée	111
Chapitre 4 - Discussion générale	
<u>Partie 1</u>	
Nature des souches Gardel et Sénégal	156
<u>Partie 2</u>	
Mécanismes de virulence et d'atténuation d' <i>Ehrlichia ruminantium</i>	159
Chapitre 5 -Conclusion et Perspectives	167
Références bibliographiques	171

I. Chapitre 1 : Introduction

SOMMAIRE

1. La cowdriose	2
1.1. Historique.....	3
1.2. Répartition géographique et risque de dispersion sur le continent américain... ..	4
1.3. Les espèces affectées	5
1.4. Les vecteurs.....	6
1.5. Symptômes et pathologies	7
1.6. Diagnostics	9
1.7. Traitement et contrôle.....	9
2. Agent causal de la cowdriose : <i>Ehrlichia ruminantium</i>	12
2.1. Morphologie et cycle de développement.....	13
2.2. Caractéristiques génétiques d' <i>E. ruminantium</i>	14
2.3. Etude de la famille multi-génique <i>Map</i> d' <i>E. ruminantium</i>	16
3. Caractérisations des interactions hôtes-pathogènes-vecteurs par des approches intégratives à haut débit.....	18
3.1. Article de revue 1: Tick-borne diseases in cattle: Applications of proteomics to develop new generation vaccines Journal of Proteomics, 2012.	19
3.2. Article de revue 2: "Omics" approaches in <i>Rickettsiales</i> : Recent progress in understanding host-pathogen-vector interactions	40
4. Objectifs de la thèse.....	85

Le centre de coopération internationale en recherche agronomique pour le développement (CIRAD) a pour vocation première, l'aide au développement agricole des pays du sud. L'unité mixte de recherche CIRAD-INRA « Contrôle des Maladies Animales Exotiques et Emergentes » au sein de laquelle j'ai effectué ma thèse, a pour ambition de contribuer par la recherche à l'amélioration du développement socio-économique des pays du Sud, en les aidant à améliorer la santé animale. Les effets attendus sont l'amélioration de la productivité animale (augmentation du revenu des éleveurs et de la disponibilité de produits animaux), la facilitation des échanges internationaux d'animaux et de produits animaux, et la protection de la santé publique en contrôlant les maladies zoonotiques. Dans ce contexte, l'UMR s'intéresse à des maladies telles que la cowdriose, maladie mortelle des ruminants qui a des impacts économiques et sociétaux très importants pour les pays du sud.

1. La cowdriose

La cowdriose est une maladie mortelle des ruminants domestiques et sauvages, elle est transmise par une tique du genre *Amblyomma*.

En Afrique, la productivité du bétail est sévèrement affectée par les tiques et les maladies qu'elles transmettent ([Musoke, et al., 1997](#)) et la cowdriose peut causer jusqu'à 90% de perte dans un troupeau, chez les races les plus sensibles.

Avec les theilérioses et les trypanosomoses, la cowdriose est considérée comme l'une des plus importantes maladies infectieuses transmises par les tiques chez les ruminants en zone tropicale.

Dans la communauté des pays sud africain, les pertes dues à la cowdriose sont estimés à 47,6 millions de dollars par an. Ces pertes sont dues à la mortalité, à la diminution de la productivité du bétail, aux coûts des acaricides et antibiotiques.

La cowdriose constitue donc un frein à l'amélioration de la productivité des systèmes d'élevage dans les pays où elle est présente ([Uilenberg, 1982](#)).

1.1. Historique

La première mention écrite de la cowdriose est due à Louis Trichardt, en Afrique du Sud. Dans son journal en date du 17 février 1838, il décrit une maladie nerveuse et mortelle des moutons ayant fait suite à une infestation massive par des tiques survenue trois semaines auparavant ([Camus, et al., 1996](#)).

En 1898, Dixon et Edington réussirent à reproduire la maladie dans des animaux sensibles par inoculation de sang d'animaux infectés. A cette époque, l'agent de la cowdriose fut considéré comme un virus du fait de son caractère filtrable et invisible au microscope.

Ce n'est qu'en 1925 que Cowdry, rickettsiologiste américain, identifia l'agent causal de la maladie. Il en décrit sa morphologie et sa localisation dans les tissus de l'hôte et du vecteur ([Cowdry, 1925](#)). Il classa la bactérie dans le genre *Rickettsia* et la nomma *Rickettsia ruminantium*. Le genre Cowdria fut créé en 1947 et le nom de la bactérie changé en *Cowdria ruminantium*. En 2001, à partir des séquences de l'ARN 16S, la bactérie fut reclassée dans le genre *Ehrlichia* et renommée *Ehrlichia ruminantium* ([Dumler, et al., 2001](#)).

1.2. Répartition géographique et risque de dispersion sur le continent américain

La cowdriose s'étend à l'ensemble de l'Afrique sub-saharienne. Elle est associée à la présence des tiques vectrices dans ces régions. La maladie existe aussi dans les îles proches du continent africain comme Madagascar, Zanzibar, La Réunion, l'île Maurice, les Comores et Sao Tomé. Dans les années 1980, elle a été mise en évidence dans les Caraïbes et fût probablement introduite dans ces îles dès le XVII^e siècle lors du commerce triangulaire avec le transport d'esclave et de bétail en provenance d'Afrique de l'Ouest. Entre les années 1960 et 1970, *Amblyomma variegatum* le vecteur principal de la cowdriose dans les Caraïbes, a progressivement infesté 16 îles depuis Sainte-Lucie et Barbade au sud jusqu'à Porto Rico au nord. A l'heure actuelle, cinq îles des Caraïbes ne sont pas infestées par les tiques du genre *Amblyomma*, il s'agit d'Anguilla, Barbuda, Montserrat, Grenade et de la Barbade. Deux autres îles, quant à elle, ne connaissent qu'une infestation mineure (Saint Vincent et Saint Croix).

Aujourd'hui, de part la présence des vecteurs capable de transmettre le pathogène et par le mouvement des oiseaux migrateurs potentiellement porteurs de tiques infectées ([Barre, et al., 1987](#)), la cowdriose constitue une menace pour le continent américain. Un tel scénario aurait un impact économique désastreux dans la région avec des pertes estimées à 792 millions de dollars par an selon l'Office international des epizooties (OIE). Un modèle mathématique basé sur les données de la biologie de la tique vectrice *Amblyomma* et sur des paramètres tels que la température et l'humidité a permis de définir les habitats favorables d'*Amblyomma variegatum* et sa possible colonisation au continent américain (figure 1).

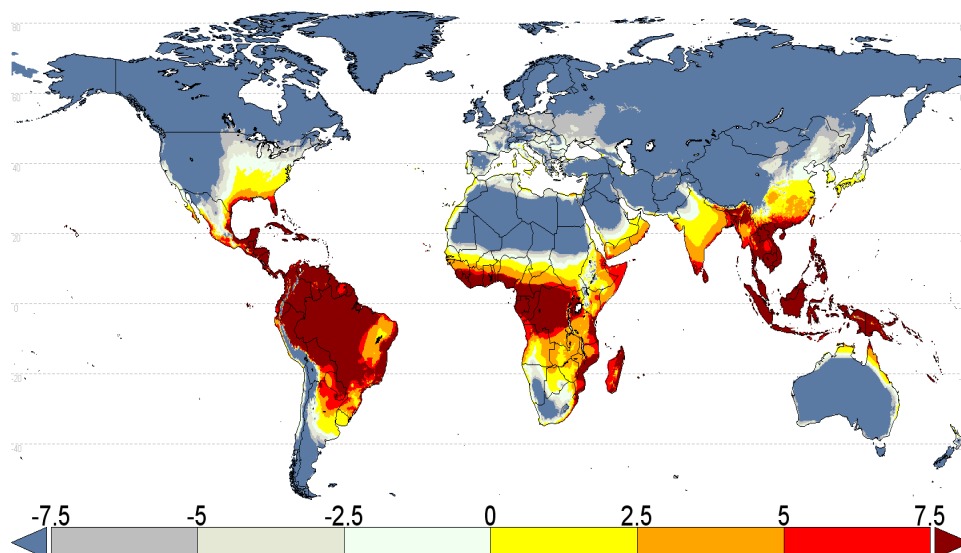


Figure 1 : Habitat favorable de la tique *Amblyomma variegatum* selon le modèle mathématique développé par T. Porphyre au sein de l'UMR (incluant la température et l'humidité). L'échelle correspond à la probabilité de maintien de la population en cas d'introduction ou de présence (rouge = forte probabilité, gris = faible probabilité).

1.3. Les espèces affectées

Tous les bovidés domestiques sont sensibles à la cowdriose : bovins, caprins et ovins, mais l'expression clinique est très variable selon les espèces ou les races, allant du portage asymptomatique au tableau clinique fatal ([Norval, et al., 1990](#)). Il existe une sensibilité à la maladie qui est dépendante de l'espèce : les caprins sont les plus sensibles puis les ovins, les bovins étant les plus résistants. Les ruminants sauvages (bovidés, cervidés) peuvent être aussi infectés mais le principal réservoir sauvage serait constitué par les buffles (*Syncerus caffer*) qui, après infection, peuvent rester porteurs du pathogène durant plusieurs mois ([Euzeby & Tindall, 2001](#)). D'autres vertébrés tels que la tortue léopard, la pintade et le lièvre des broussailles présentent une rickettsiémie après infection expérimentale

([Oberem & Bezuidenhout, 1987](#)). Deux espèces de rongeurs sauvages se sont aussi montrées sensibles expérimentalement et, en 1971, Du Plessis et Kümm ont décrit un premier isolat d'*E. ruminantium* pathogène pour la souris de laboratoire ([du Plessis & Kumm, 1971](#)). Des souches d'*E. ruminantium* telles que Welgevonden et Ball3, ont depuis pu être isolées chez la souris.

Chez l'homme, un risque potentiel d'infection peut exister, car *E. ruminantium* est cultivable en cellules endothéliales humaines (HUVEC : cellules humaines de cordon ombilical) ([Totte, et al., 1993](#)), cependant la cowdriose n'est pas reconnue comme étant une zoonose ([Kelly, et al., 1992](#)). En 2005, trois cas humains d'infection par *E. ruminantium* ont été rapportés en Afrique du Sud ([Allsopp, et al., 2005](#)) avec détection du pathogène par PCR sur les prélèvements, cependant il n'a pas été prouvé de lien entre la mortalité des patients et la présence du pathogène.

1.4. Les vecteurs

A ce jour, 13 espèces de tiques appartenant toutes au genre *Amblyomma* sont capables de transmettre *E. ruminantium* dans les conditions naturelles ou lors d'infections expérimentales. *A. variegatum* (vecteur principal) et *A. hebraeum* sont les deux espèces vectrices majeures. Dans les caraïbes, *A. variegatum*, est la seule espèce présente. Des espèces de tiques présentes sur le continent américain ont pu être infectées expérimentalement et représentent des vecteurs potentiels de la cowdriose : *A. cajennense*, *A. maculatum*, *A. dissimile* et *A. americanum*.

Les tiques du genre *Amblyomma* sont des tiques trixènes (3 hôtes), chacun des trois stades de développement (larve, nymphe, adulte) prenant un repas de sang sur un hôte différent. L'infection des tiques a lieu aux stades larvaire ou nymphal par gorgement sur un animal malade ou porteur asymptomatique et conduit, après la mue, à l'obtention de nymphes ou d'adultes infectées. La transmission de la bactérie à un nouveau ruminant

ne survient qu'après 1 à 3 jours de gorgement pour les nymphes et 2 à 4 jours pour les adultes ([Bezuidenhout, 1987](#), [Camus & Barre, 1995](#)). La transmission transstadiale a été montrée par contre il n'existe pas de transmission transovarienne du pathogène.



Figure 2 : Tiques de l'espèce *Amblyomma variegatum* en gorgement sur une chèvre (Cliché du CIRAD)

1.5. Symptômes et pathologies

a) Symptômes

Le tableau clinique de la cowdriose peut varier d'une maladie suraigue à bénigne en fonction de l'espèce, de la race, de l'âge et du statut immunitaire de l'animal, ainsi que de la souche et de la quantité de bactéries inoculées.

L'infection des petits ruminants conduit en général à des formes suraiguës ou aiguës de cowdriose. Les signes majeurs caractéristiques de ces formes graves sont une hyperthermie soudaine et très élevée (souvent supérieure à

41°C), de la prostration due à l'état fébrile avec la tête portée basse, un poil terne et légèrement ébouriffé, une détresse respiratoire et, finalement, l'apparition de symptômes nerveux. L'évolution dure de 1 à 6 jours dans les formes aiguës et la mort est l'issue quasi certaine, dès lors que les symptômes nerveux sont apparus.

Dans les formes suraiguës, l'animal est retrouvé mort sans avoir montré de signes cliniques, ou il meurt en quelques heures dans une crise de convulsions après avoir développé une hyperthermie élevée.

Les formes subaiguës peuvent être observées sur des animaux développant un certain degré de résistance génétique ou une immunité acquise partielle, ou sur des animaux infectés avec une faible quantité de bactérie ou avec une souche de faible virulence. Les signes cliniques sont identiques mais atténués et l'évolution est plus longue. Après 7 à 10 jours, l'animal amaigri peut mourir. La guérison est possible mais la convalescence est en général longue, l'animal pouvant rester amaigri pendant plusieurs semaines.

Toutes les formes décrites chez les petits ruminants peuvent être observées chez les bovins. Les symptômes nerveux sont moins constants que chez les petits ruminants mais les formes digestives, caractérisées par une diarrhée fétide et profuse, sont très fréquentes.

b) Pathologies

Le tableau lésionnel varie en fonction de la forme de la maladie, mais se caractérise globalement par des épanchements de liquides dans les grandes cavités : hydrothorax, hydropéritoine, hydropéricarde ([Prozesky, 1987](#), [Prozesky, 1987](#)). L'hydropéricarde est présent dans la quasi-totalité des cas chez les petits ruminants. Cette lésion est à l'origine du nom anglais de la maladie « Heartwater ». Des lésions moins caractéristiques peuvent également être observées : splénomégalie, hypertrophie et œdème des nœuds lymphatiques mésentériques et préscapulaires, congestion du foie, pétéchies et petites hémorragies sur le myocarde, le péricarde et l'intestin. Le cerveau ne présente pas de lésions macroscopiques évidentes, excepté une congestion et un œdème des méninges. Les lésions histologiques

observées, oedème pulmonaire et péricardique, résultent d'une augmentation de la perméabilité vasculaire (processus inflammatoire aigu), le liquide infiltrant contenant en majorité des macrophages. De façon générale, il n'y a pas de corrélation entre la sévérité de la maladie et des lésions, et le nombre de bactéries présents dans les tissus infectés.

1.6. Diagnostics

L'observation des symptômes cliniques tels qu'une hyperthermie élevée, des signes nerveux, une détresse respiratoire et une évolution vers la mort en 2 à 5 jours, doivent évoquer la cowdriose en zone infestée de tiques du genre *Amblyomma*. Cependant, ces symptômes peuvent être confondus avec d'autres maladies bactériennes ([Uilenberg, 1983](#)).

Classiquement, la confirmation du diagnostic microscopique de la cowdriose après la mort d'un animal, se fait par la recherche de colonies d'*E. ruminantium* dans les cellules endothéliales des vaisseaux capillaires cérébraux à partir d'un frottis de cerveau coloré au GIEMSA ou au RAL555. Il existe des tests ELISA mettant en évidence les anticorps anti MAP1 cependant ces tests ne sont pas utilisés pour du diagnostic de cowdriose. Le diagnostic moléculaire ciblant le gène *pCS20* est utilisé sur des cas cliniques de suspicion de cowdriose.

1.7. Traitement et contrôle

a) Traitement par des antibiotiques

L'oxytétracycline administrée par voie intramusculaire ou intraveineuse est particulièrement efficace et constitue le traitement le plus utilisé pour traiter les infections à *E. ruminantium* ([Van Amstel & Oberem, 1987](#)). Les formulations longue-action sont également très efficaces. Le succès thérapeutique dépend essentiellement de la précocité du traitement. Une à trois injections d'oxytétracycline administrées en fonction du retour ou non

à la normale de la température rectale, suffise à guérir l'animal. Administré en phase d'hyperthermie, le traitement conduit presque toujours à la guérison avec induction d'une immunité forte. Lorsque les symptômes nerveux apparaissent, la mort est l'issue la plus fréquente.

b) *Lutte contre le vecteur*

La lutte antivectorielle est basée sur l'utilisation d'acaricides. Ce moyen de lutte est coûteux et nocif pour l'environnement. De plus, il ne permet pas d'éradiquer les tiques et nécessite un traitement fréquent.

c) *Lutte vaccinale*

La seule méthode de vaccination commerciale utilisée exclusivement en Afrique du Sud, est basée sur une infection avec un inoculum infectieux (sang, broyat de tiques) et un traitement aux tétracyclines lors de l'apparition de l'hyperthermie ([Oberem & Bezuidenhout, 1987](#)). Même s'il confère une immunité forte et de longue durée, il a de nombreux désavantages comme le respect de la chaîne du froid, l'utilisation d'un stabilat infectieux non contrôlé, un suivi journalier de la température et une mortalité possible des animaux.

Depuis, divers essais de développement de vaccins ont été réalisés. L'immunisation utilisant un vaccin inactivé à partir d'*E. ruminantium* cultivés sur des cellules endothéliales bovines a été développée par Martinez et collaborateurs (1994). Le vaccin inactivé confère une bonne protection (80 à 100% de protection) pour des animaux immunisés et infectés avec une souche homologue et une protection plus faible lors d'une infection avec une souche hétérologue ([Martinez, et al., 1994](#)). Le principal avantage du vaccin inactivé est la possibilité de fournir un vaccin multivalent avec un cocktail de plusieurs souches à condition qu'elles soient isolées et cultivées *in vitro*. Il a été montré récemment qu'un vaccin incluant une souche de référence Gardel et une souche locale permettait une bonne protection de moutons sensibles à la cowdriose lors d'essais sur le terrain au Burkina Faso ([Adakal, et al., 2010](#)). Le processus industriel de production des antigènes en bioréacteur a été mis au point ([Marcelino, et al., 2007](#)) et il

existe maintenant un vaccin « prêt à l'emploi émulsionné en adjuvant huileux ISA70 dont l'efficacité a été montrée en conditions contrôlées contre une infection avec une souche homologue à la souche vaccinale.

Des souches d'*E. ruminantium* ont été atténuées *in vitro*, offrant ainsi la possibilité de l'utilisation de vaccins vivants avec des souches atténuées. Il existe actuellement trois souches atténuées. Les souches Gardel et Sénégal ont pu être atténuées après de multiples passages sur cellules endothéliales (200 et 15 passages respectivement) ([Jongejan, et al., 1993](#)). La souche Welgevonden a aussi pu être atténuée *in vitro* dans des macrophages canins ([Zweygarth, et al., 2005](#)). La méthode de vaccination par inoculation de souches atténuées a montré son efficacité dans le cas d'infections homologues ([Faburay, et al., 2007](#)). Le vaccin atténué comme le vaccin inactivé confère une protection plus faible lors d'infection avec des souches hétérologues des souches vaccinales. De plus, il n'existe que peu de souches atténuées auquel s'ajoute deux contraintes concernant le respect de la chaîne du froid et le risque potentiel de réversion de l'atténuation. D'autres travaux ont été réalisés sur les protéines de la famille multigénique *map* (Major antigen protein). Un plasmide exprimant la protéine MAP1 a été produit mais les essais sur les ruminants n'ont pas été concluants surtout en ce qui concerne la protection en challenge hétérologue ([Nyika, et al., 1998](#), [Nyika, et al., 2002](#))

L'utilisation de ces vaccins inactivés ou atténués permet d'obtenir de bons niveaux de protection dans le cas d'une infection avec une souche homologue mais, au sein de l'espèce *E. ruminantium*, il y a une grande diversité des souches ce qui constitue un frein à l'élaboration d'un vaccin efficace.

L'élaboration d'un vaccin recombinant utilisant des gènes ou antigènes ciblant l'ensemble des souches d'*E. ruminantium* et induisant une protection contre toutes ces souches semble être la meilleure solution. Sur ce principe, deux souches d'*E. ruminantium* possédant des phénotypes différents (absence de protection croisée) ont été séquencées dans le but de trouver en

autre des gènes candidats à l'élaboration de vaccin recombinant, les antigènes protecteurs étant probablement polymorphes. Les deux souches sont: *E. ruminantium* Gardel et Welgevonden ([Collins, et al., 2005](#)); ([Frutos, et al., 2006](#)). Récemment, trois gènes polymorphiques pour les souches d'*E. ruminantium* ont été sélectionnés. Il s'agit de *CpG1*, *CpG2* et *CpG3* (Cowdria polymorphic genes). Ces gènes ont été testés comme vaccin ADN plasmidique et aucune efficacité n'a été montrée. Par contre, lorsqu'une pré-immunisation avec la protéine recombinante CPG1 est effectuée avant l'immunisation plasmidique *CpG1* il y a une protection des animaux contre une infection homologe ([Pretorius, et al., 2010](#)). Cependant, étant donné la nature polymorphe de *CpG1*, une protection efficace sur le terrain ne semble possible que si un mélange de *CpG1* de souches différentes est inclus dans le vaccin. Jusqu'à présent, ces vaccins recombinants utilisant une pré vaccination ADN puis un rappel avec protéines recombinantes sont difficilement applicables sur le terrain à cause des moyens de délivrance utilisés (« gene gun »).

2. Agent causal de la cowdriose : *Ehrlichia ruminantium*

Ehrlichia ruminantium est une bactérie intracellulaire obligatoire classée dans la subdivision des *α-Proteobactéries*. Elle appartient à l'ordre des *Rickettsiales*, dans la famille des *Anaplasmataceae* qui regroupe aussi les genres *Anaplasma*, *Ehrlichia*, *Neorickettsia* et *Wolbachia*. *E. ruminantium* est capable d'infecter les monocytes, les neutrophiles et les cellules endothéliales des ruminants. Les membres de la famille des *Anaplasmataceae* sont des bactéries Gram négatif mais différent, de la majorité des bactéries Gram négatif par l'absence de production d'endotoxines et par leur transmission essentiellement vectorielle.

2.1. Morphologie et cycle de développement

E. ruminantium, est une bactérie aérobie de 0,2 à 2,7 μm , polymorphe, elle se présente sous une forme coccoïde ou ellipsoïdale. Les observations au microscope électronique ont permis de montrer l'existence de trois stades de développement ([Jongejan, et al., 1991](#)):

- le stade corps élémentaire (CE) ou corps initial dense (dense-cored form), de forme ovoïde, de petite taille (0,6 à 1,2 μm) et présentant des filaments nucléiques condensés (Figure 3). Il est limité par une double membrane, et constitue la forme libre et infectieuse de la bactérie, adhère et pénètre dans la cellule hôte par un processus qui ressemble à la phagocytose.
- le stade corps réticulé (CR) présentant une forme pléomorphe de 1,2 μm à 2,8 μm et des filaments nucléiques dispersés ([van Vliet, et al., 1992](#)). Les corps réticulés sont considérés comme des formes végétatives non infectieuses.
- le stade corps intermédiaire : de forme intermédiaire entre le corps élémentaire et le corps réticulé, de taille de 0,9 μm à 2,2 μm issue de la modification des corps réticulés et ayant un ADN plus condensé.

Une fois à l'intérieur de la cellule hôte, les CE deviennent des CR qui se multiplient au sein d'une vacuole cytoplasmique bordée d'une double membrane dérivant de la cellule hôte. Dans la vacuole, les organismes se divisent par fission binaire jusqu'à former une large colonie dénommée *morula* (figure 3). Après environ 4 à 6 jours de développement *in vitro*, la lyse de la cellule hôte conduit à la libération de centaines de CE qui peuvent initier de nouveaux cycles infectieux.

Chez le vecteur, *E. ruminantium* se développe dans les cellules épithéliales de l'intestin. Elle gagne ensuite l'hémolymphe puis *in fine* les glandes salivaires (Figure 3). Lors du repas sanguin, les bactéries présentes dans la salive de la tique peuvent ainsi infecter d'autres hôtes.

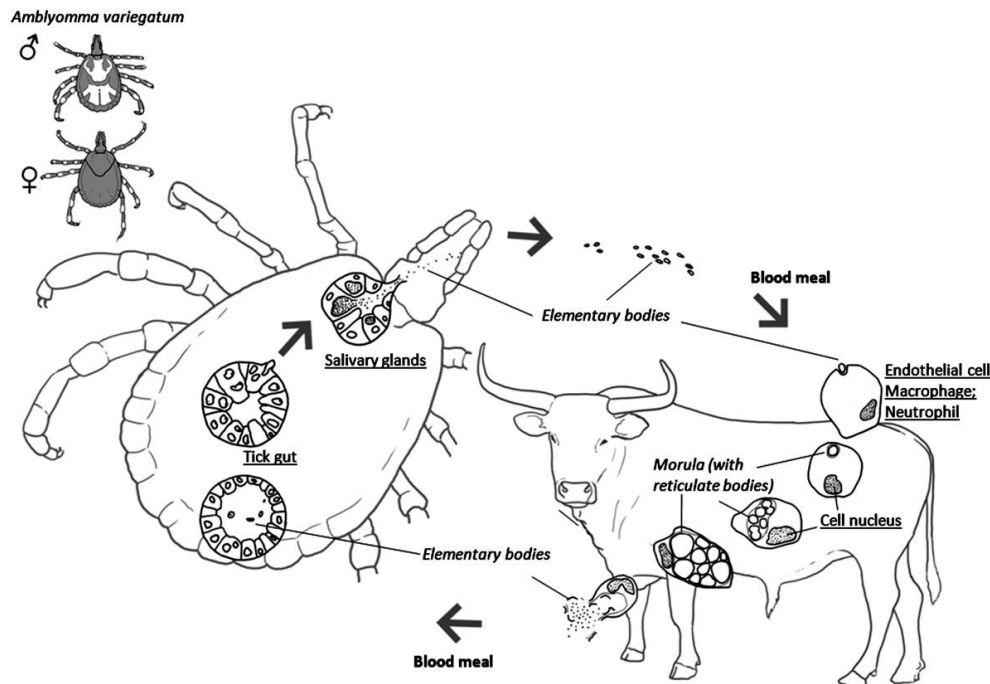


Figure 3 : Schéma du cycle de développement d'*E. ruminantium* chez son hôte et son vecteur ([Marcelino, et al., 2012](#))

2.2. Caractéristiques génétiques d'*E. ruminantium*

Actuellement, trois souches d'*E. ruminantium* ont été entièrement séquencées et publiées ([Collins, et al., 2005](#)); [Frutos, et al., 2006](#)). Il s'agit de la souche Welgevonden (Erwo) isolée en Afrique du Sud ([Du Plessis, 1985](#)), d'un isolat de cette souche maintenu dans un environnement cellulaire différent en Guadeloupe, la souche Erwe et de la souche Gardel (Erga) isolée en Guadeloupe ([Uilenberg, et al., 1985](#), [Frutos, et al., 2006](#)). Le génome d'*E. ruminantium* présente une taille d'environ 1,5 mégabases avec un nombre de CDS compris entre 920 et 958 suivant les souches (Tableau 1). Le contenu en GC % est relativement faible avec une valeur moyenne de 27,5%. L'analyse en génomique comparative de ces souches a révélé l'existence de 3 régions RAD (regions of accumulated differences) dans lesquelles des troncations de gènes ont été identifiées, de même qu'un petit nombre de gènes spécifiques de souches ([Frutos, et al., 2006](#)).

Contrairement aux autres *Rickettsiales*, *E. ruminantium* présente un faible pourcentage de génome codant (64%), qui est associé à la présence de longs espaces inter-géniques. A l'intérieur de ces espaces inter-géniques, il existe un grand nombre de séquences répétées en tandem (tandem repeated sequences) ([Collins, et al., 2005](#), [Frutos, et al., 2006](#), [Frutos, et al., 2007](#)).

A ce jour, trois nouvelles souches ont été séquencées et sont en cours d'annotation. Il s'agit de la souche Gardel atténuée et des souches Sénégal virulente et atténuée dont les données ont été utilisées dans le cadre de cette thèse.

Caractéristiques	Génome		
	Erga ^a	Erwe ^a	Erwo ^b
Taille du génome (pb)	1499920	1512977	1516355
% G+C	27.51 %	27.48 %	27.5 %
Nombre de CDS prédits	950	958	920
% de génome codant	64.4 %	63.8 %	63.1 %
Taille moyenne des CDS	1007	998	1032
Nombre d'ARNt	36	36	36
Nombre d'ARNr	3	3	3

a : d'après Frutos *et al.*, 2006

b : d'après Collins *et al.*, 2005

Tableau 1 : Caractéristiques générales des génomes d'*E. ruminantium* Gardel (Erga), Welgevonden (Erwe et Erwo)

Le premier gène cloné et décrit chez *E. ruminantium* a été *map1* (*major antigenic protein*, codant pour une protéine antigénique majeure de surface). Le séquençage du gène a révélé qu'il est constitué de parties conservées et

de 3 régions hypervariables; le polymorphisme du gène *map1* est mis à profit pour le typage moléculaire des souches. *E. ruminantium* présente une grande diversité génétique. En effet plus d'une quarantaine de souches est aujourd'hui dénombrée et 9 génotypes *map1* différents ont été découverts en Guadeloupe. Cette diversité génétique au sein de l'espèce *E. ruminantium* rend le traitement vaccinal difficile.

Afin de mieux caractériser cette diversité génétique, deux types d'approches de typage moléculaire ont été développés au sein de l'équipe. L'une de ces techniques est basée sur le polymorphisme généré par la présence dans le génome des séquences répétées en tandem en nombre variable ou VNTR (Variable Number Tandem Repeats), MLVA (Multi Locus VNTR Analysis). Huit VNTR ont été ainsi validés sur 17 échantillons d'*E. ruminantium* extraits de tiques ou d'organes d'animaux infectés d'origines géographiques différentes ([Pilet, et al., 2012](#)). Cette technique pourra être utilisée à grande échelle pour le diagnostic moléculaire et le typage d'échantillons provenant de tiques infectées ou d'organes d'animaux malades. L'autre technique de typage, est basée sur le polymorphisme de gènes dit de « ménages » codant pour des fonctions indispensables à la vie cellulaire. Ces gènes sont faiblement soumis aux variations génétiques et permettent ainsi de mettre en évidence des phénomènes de polymorphismes ancestraux. La technique MLST est essentiellement utilisée pour des études phylogéniques et phylogéographiques.

2.3. Etude de la famille multi-génique *Map* d'*E. ruminantium*

La protéine membranaire MAP1 (Major Anigenic Protein 1) a été largement étudiée du fait de la mise en évidence d'anticorps majoritaires dirigés contre cette protéine lors de l'infection chez l'hôte ([van Vliet, et al., 1994](#)). Elle appartient à une famille multigénique *map* contenant 16 paralogues de gènes codant pour des protéines membranaires externes (Outer Membrane Protein, OMP). Même s'il a été prouvé que MAP1

n'est pas une protéine induisant une réponse immunitaire protectrice, cette famille multigénique pourrait avoir un intérêt dans une approche vaccinale.

Une première étude de l'expression des gènes de cette famille multigénique a été effectuée sur *E. ruminantium* cultivée dans des cellules endothéliales bovines et dans deux lignées différentes de cellules de tiques ([Van Heerden, et al., 2004](#), [Bekker, et al., 2005](#)). Cette étude montre que si tous les gènes *map* sont exprimés dans les cellules endothéliales, seules certains sont exprimés dans les cellules de tiques ([Bekker, et al., 2005](#)). De plus, le pool de gènes exprimés dans les cellules de tiques varie suivant l'espèce et l'origine des cellules. Le profil d'expression des gènes est cependant conservé entre les 3 souches d'*E. ruminantium* testées (Gardel, Welgevonden, Senegal). Une étude d'expression des gènes de la famille *map* a été réalisée *in vivo* sur des glandes salivaires et l'intestin de tiques non gorgées et gorgées. Dans les tiques non gorgées les transcrits de *map1-1* sont détectés dans l'intestin et aucun transcrit n'est détecté dans les glandes salivaires. Après gorgement, *map1-1* est plus abondant dans l'intestin alors que les transcrits de *map1* sont présents en grande quantité dans les glandes salivaires ([Postigo, et al., 2007](#)). Il existe des modifications d'expression des *map* en fonction des tissus et du stade de gorgement qui suggèrent qu'elles jouent un rôle important dans le développement de la bactérie.

En 2008, une stratégie similaire a été effectuée pour l'étude de l'expression des protéines MAP ([Postigo, et al., 2008](#)). Dans cette étude, les auteurs ont comparé l'expression des protéines de la famille MAP dans les bactéries cultivées en cellules endothéliales ou en cellules de tiques. La protéine MAP1 était majoritairement détectée dans les cellules endothéliales alors que, dans les cellules de tiques, la protéine majoritairement détectée était MAP1-1 ([Postigo, et al., 2008](#)).

L'expression différentielle des gènes et des protéines de la famille multigénique *map* semblent être liée étroitement au type cellulaire. Cette famille multigénique semble donc intervenir dans les interactions ou les adaptations d'*E. ruminantium* à l'hôte et au vecteur.

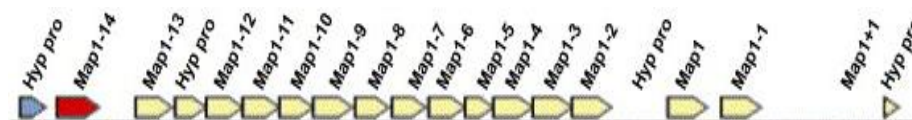


Figure 4 : Cluster des gènes de la famille *Map* d'*E. Ruminantium*

3. Caractérisations des interactions hôtes-pathogènes-vecteurs par des approches intégratives à haut débit

Les maladies provoquées par les pathogènes transmises par les tiques dont les *Rickettsiales* ont un impact économique important et des conséquences graves, sur les animaux et l'homme. Elles nécessitent donc le développement de stratégies de lutte efficace dont l'élaboration de vaccins de seconde génération. Cependant, le caractère intracellulaire obligatoire de ces pathogènes, représente un obstacle important dans l'étude de leur pathogénèse et l'élaboration de vaccins efficaces.

Des approches haut débit « omics » telles que la génomique comparative, la transcriptomique, ou encore la protéomique sont aujourd'hui utilisées dans le but d'élucider les mécanismes de pathogénicité de ces agents pathogènes. Nous présentons sous la forme de deux articles de revue les dernières études utilisant les technologies « omics » pour l'étude d'une part de pathogènes transmis par les tiques et d'autres part pour l'étude des *Rickettsiales*. Dans ces 2 revues, les études sur notre pathogène d'intérêt *Ehrlichia ruminantium* sont présentées. Dans le premier article, sont répertoriés principalement les résultats de protéomique sur les pathogènes provoquant les maladies les plus importantes chez les ruminants : theileriose, babesiose, anaplasmose et cowdriose. Dans le

second article de revue, nous nous sommes intéressés aux études récentes menées chez les *Rickettsiales* par les approches « omics » et permettant de comprendre les relations entre les pathogènes, leurs hôtes et leurs vecteurs mais surtout d'analyser les manques et les nouvelles perspectives de recherches. Ces dernières années, les améliorations concernant les conditions de culture *in vitro*, l'isolement de souches, mais aussi la disponibilité des données de séquençage de plusieurs pathogènes, de leur hôtes et vecteurs ont permis des avancées significatives sur la connaissance des mécanismes qui interviennent dans la pathogenèse et la virulence de ces organismes.

3.1. Article de revue 1: Tick-borne diseases in cattle: Applications of proteomics to develop new generation vaccines Journal of Proteomics, 2012

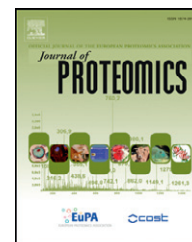
Les theilerioses et les babesioses sont causées respectivement par les parasites protozoaires *Theileria parva*, *Theileria annulata*, *Babesia bovis* et *Babesia bigemina* tandis que la cowdriose et l'anaplasmose bovine sont dues respectivement à *E. ruminantium* et *A. marginale*, deux bactéries intracellulaires strictes de la famille des *Anaplasmataceae*. L'une des stratégies de lutte contre ces maladies, est le contrôle du vecteur par l'utilisation d'acaracides entraînant ainsi des conséquences désastreuses pour l'environnement et le développement de résistance des tiques à ces produits. Comme pour d'autres maladies causées par des pathogènes intracellulaires, l'autre stratégie de lutte envisagée contre ces maladies, est le développement de vaccins ciblant le pathogène lui-même ou ses déterminants nécessaires à l'infection.

Les premières stratégies vaccinales développées reposaient sur l'utilisation d'inoculum vivant (sang, broyat) mais, ces vaccins présentent d'importants inconvénients tels que le respect de la chaîne du froid ou encore la transmission potentielle d'autres pathogènes contenu dans l'inoculum. Le développement de nouvelles stratégies vaccinales plus efficaces est donc

nécessaire dans la lutte contre ces maladies transmises par les tiques. Le développement de ces nouvelles stratégies passent par une connaissance approfondie du pathogène mais aussi des mécanismes lui permettant d'infecter et de se développer chez son hôte et/ou son vecteur. Pour les 4 pathogènes, les études d'expression des protéines ont été réalisées dans les cellules de l'hôte et de la tique. Pour *Anaplasma marginale*, un intérêt particulier a été porté sur les protéines de surface sachant qu'elles induisent une immunité protectrice chez les ruminants. Vingt quatre protéines immuno-dominantes ont été identifiées incluant les protéines de membranes externes Msp2, Msp3 et Msp5 mais aussi des protéines du système de sécrétion de type IV (SST4), les protéines VirB9 et VirB10. Des protéines de surface externe comprenant Omp1, Omp7-9, Msp1a, Msp2, Msp3, Msp4, OpAG2 and Am779 ont été utilisées pour vacciner des animaux et sont capables d'induire une réponse immunitaire protectrice. Ceci montre l'importance de l'étude des protéines membranaires dans la recherche de vaccin contre *Anaplasma marginale*. La comparaison de l'expression des protéines de surface d'*Anaplasma* dans les cellules hôtes et les cellules de tiques a mis en évidence une plus grande complexité dans les cellules hôtes avec 15 protéines exprimées *vs* 5 chez la tique. Quatre protéines sont communes entre cellules de tiques et hôte et Am778 étant la protéine unique d'*Anaplasma* trouvée chez la tique. Il existe une modification de l'expression des protéines du pathogène entre l'hôte et la tique. Ces protéines pourraient être importantes dans la transmission de l'hôte au vecteur.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Review

Tick-borne diseases in cattle: Applications of proteomics to develop new generation vaccines[☆]

Isabel Marcelino^{a,b,*,1}, André Martinho de Almeida^{b,c,d,1}, Miguel Ventosa^{a,b},
Ludovic Pruneau^e, Damien F. Meyer^e, Dominique Martinez^f, Thierry Lefrançois^e,
Nathalie Vachiéry^e, Ana Varela Coelho^b

^aIBET, Apartado 12, 2780-901 Oeiras, Portugal^bInstituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa (ITQB-UNL), Av. da República, 2780-157 Oeiras, Portugal^cIICT/CVZ — Instituto de Investigação Científica Tropical, Centro de Veterinária e Zootecnia, Av. Univ. Técnica, 1300-477 Lisboa, Portugal^dCIISA — Centro Interdisciplinar de Investigação em Sanidade Animal, Fac. Med. Veterinária, Av. Univ. Técnica, 1300-477 Lisboa, Portugal^eCIRAD, UMR CMAEE, F-97170 Petit-Bourg, Guadeloupe, France^fCIRAD, UMR CMAEE, F-34398 Montpellier, France

ARTICLE INFO

Article history:

Received 2 January 2012

Accepted 15 March 2012

Available online 26 March 2012

Keywords:

Tick-borne diseases

Cattle

Vaccines

Proteomics

Transcriptomics

Systems biology

ABSTRACT

Tick-borne diseases (TBDs) affect 80% of the world's cattle population, hampering livestock production throughout the world. Livestock industry is important to rural populations not only as food supply, but also as a source of income. Tick control is usually achieved by using acaricides which are expensive, deleterious to the environment and can induce chemical resistance of vectors; the development of more effective and sustainable control methods is therefore required.

Theileriosis, babesiosis, anaplasmosis and heartwater are the most important TBDs in cattle. Immunization strategies are currently available but with variable efficacy. To develop a new generation of vaccines which are more efficient, cheaper and safer, it is first necessary to better understand the mechanisms by which these parasites are transmitted, multiply and cause disease; this becomes especially difficult due to their complex life cycles, *in vitro* culture conditions and the lack of genetic tools to manipulate them.

Proteomics and other complementary post-genomic tools such as transcriptomics and metabolomics in a systems biology context are becoming key tools to increase knowledge on the biology of infectious diseases. Herein, we present an overview of the so called “Omics” studies currently available on these tick-borne pathogens, giving emphasis to proteomics and how it may help to discover new vaccine candidates to control TBDs.

This article is part of a Special Issue entitled: Farm animal proteomics.

© 2012 Elsevier B.V. All rights reserved.

[☆] This article is part of a Special Issue entitled: Farm animal proteomics.

* Corresponding author. Tel.: + 351 21 4469452.

E-mail address: marcelino@itqb.unl.pt (I. Marcelino).

¹ These authors contributed equally to this work.

Contents

1. Control of tick-borne diseases in cattle — current situation	4233
2. Studying host–vector–pathogen interactions: key issues on biological samples	4236
3. Proteomic studies on tick-borne pathogens of cattle	4237
3.1. <i>Theileria parva</i> and <i>Theileria annulata</i>	4237
3.1.1. Proteomics studies in host cells	4237
3.1.2. Proteomics studies in tick cells	4238
3.2. <i>Babesia bovis</i> and <i>Babesia bigemina</i>	4238
3.2.1. Proteomic studies in host cells	4238
3.2.2. Proteomic studies in tick cells	4239
3.3. <i>Anaplasma marginale</i>	4241
3.3.1. Proteomic studies in host cells	4241
3.3.2. Proteomic studies in tick cells	4242
3.4. <i>Ehrlichia ruminantium</i>	4243
3.4.1. Proteomic studies in host and tick cells.	4243
4. Complementing proteomics with transcriptomics analyses	4243
4.1. <i>Theileria</i> spp.	4244
4.2. <i>Babesia</i> spp.	4244
4.3. <i>Anaplasma marginale</i>	4245
4.4. <i>Ehrlichia ruminantium</i>	4245
5. Metabolomics — providing additional insights in pathogen virulence and survival	4246
6. System biology: integrating “Omics” data to improve vaccine development	4246
7. Conclusion	4247
Acknowledgments.	4247
References	4247

1. Control of tick-borne diseases in cattle — current situation

Tick-borne diseases (TBDs) pose a major constraint of live-stock production and have considerable economic impact to rural people affecting not only their food supply, but also their daily income and other agricultural activities [1]. In 1997, the annual global losses associated to ticks and TBDs in cattle was estimated to amount between US\$ 13.9 billion and US\$ 18.7 billion [2]. Four groups of TBDs are of importance to the livestock production: theileriosis, babesiosis, anaplasmosis and heartwater (also called cowdriosis), posing major health and management problems of cattle and small ruminants in affected areas. In cattle, East Coast fever (ECF), tropical theileriosis and babesiosis are caused by protozoan parasites (*Theileria parva*, *Theileria annulata* and *Babesia bovis* or *Babesia bigemina*, respectively), whereas heartwater and anaplasmosis are caused by the Rickettsiales *Ehrlichia ruminantium* and *Anaplasma marginale*, respectively. In endemic areas, indigenous cattle have developed resistance to ticks and to tick-borne pathogens (TBPs) [1]. However, susceptibility of exotic breeds presents a major obstacle to the improvement of cattle production.

Chemical tick control, treatment of animals, chemoprophylaxis and vaccination are among the measures available to limit losses incurred by TBDs. However, chemical control is limited owing to selection of acaricide-resistant ticks; in addition, residues in meat and milk have raised public health concerns. Intensive acaricide use also interferes with enzootic stability, rendering animals susceptible to the diseases.

Chemoprophylaxis can be effective, but only for a short period. Therefore, more effective and sustainable integrated control methods such as vaccines should be developed to control TBDs [1].

Veterinary vaccines are the most economical and sustainable method to prevent and control infectious diseases in animals, to improve animal welfare and decrease the cost of animal production [3]. Additionally, the implementation of mass vaccination programs for animals can significantly contribute to reduce the consumption of different veterinary drugs, therefore preventing the emergence of resistance of microorganisms or parasites and reducing the burden of diseases, including zoonoses. Thus, besides improving the animal health sector itself, veterinary vaccines can enhance public health [4]. The advances in vaccine research for theileriosis, babesiosis, anaplasmosis and heartwater are presented below and summarized in Table 1.

Cattle are particularly vulnerable to *T. parva* and *T. annulata* infections, as they induce the transformation of T-lymphocytes and macrophages, respectively, leading to uncontrolled proliferation and metastasis of these host cells [5]. *T. parva* occurs in Eastern and Southern Africa while *T. annulata* is found in North Africa, Southern Europe and Asia; these two parasites are not usually found in the same regions. The only commercially available vaccine against *T. parva* is an infection-and-treatment method which consists of a live sporozoite challenge together with simultaneous treatment of the resultant infection with oxytetracycline; this vaccine is produced in International Livestock Research Institute (ILRI)'s Nairobi laboratories. Either sporozoites or schizonts can be used to produce live vaccines for protection against tropical

theileriosis [6]. However, the only commercialized *T. annulata* vaccine is based on attenuated shizonts produced in cell culture (Rakshavac-T®, National Dairy Development Board, India) [1]. Live sporozoites from infected ticks have been tested experimentally as a potential vaccine for tropical theileriosis but have not yet been applied in field vaccination campaigns.

Infection with the apicomplexan hemoprotozoan *B. bovis* and *B. bigemina* results in the destruction of erythrocytes (causing severe anemia) which poses serious limitations to cattle development in tropical and subtropical regions of the world. Although there has been some progress in the development of vaccines for babesiosis (Table 1), no effective and safe vaccine is currently commercially available [7]. So far the production of live anti-*Babesia* vaccines is available but it requires the supply of fresh bovine erythrocytes and serum from specific donors; these bovine donors are maintained tick-free and under highly controlled conditions to ensure that no other blood transmissible infectious agents are present.

A. marginale is the most prevalent TBP of cattle with a world-wide distribution. Acute disease manifests with anemia, weight loss, and often, death. In animals that survive acute disease, *A. marginale* establishes life-long persistent infection [8]. In susceptible animals, the control of anaplasmosis using live vaccines was initiated in the early 1900s and these continue to be the vaccines of choice in many parts of the world. These vaccines are based on erythrocytes from splenectomized calves experimentally inoculated with defined isolates of *A. marginale* or *A. centrale*. Although a single injection is enough to induce lifelong protective immunity in cattle, these blood-derived vaccines can be contaminated with bovine cells or other pathogens that frequently cause persistent infections in cattle [9]. To overcome this issue, inactivated vaccines were developed based on *A. marginale* antigen partially purified from bovine erythrocytes [9]. These blood-derived killed vaccines reduced clinical anaplasmosis but are expensive to produce, difficult to standardize, and often not cross-protective in widely separated geographic areas with different endemic *A. marginale* isolates. *A. marginale* harvested from cell culture are infective for both cattle and ticks and, when used as antigen, conferred partial protection to immunized cattle [10]. Amongst some candidate antigen as subunit vaccines, the major surface proteins 1 (MSP 1) known to be involved in adhesion to host and tick cell induced partial protective immunity in cattle [11]. No *A. marginale* live attenuated vaccine is available since attenuation of *A. marginale* has not been achieved. Despite all these efforts, there is currently no widely accepted vaccine for anaplasmosis [8].

Heartwater, caused by the Rickettsiales *Ehrlichia* (formerly known as *Cowdria*) *ruminantium*, is one of the most important diseases of livestock in Africa. It is particularly serious in non-indigenous livestock that are moved into affected areas causing mortality rates up to 90%. It has long been recognized that animals which recover from heartwater acquire immunity to the disease. Based on this, numerous attempts at producing a vaccine have been made [12]. Early trials were made prior to 1926 and included the inoculation of bile, the use of hyperimmune serum, the inoculation of sub-lethal doses, the attenuation of the heartwater agent or its inactivation with formalin [13]. Nevertheless all these attempts failed. Nowadays, four vaccine strategies against heartwater have

been developed; the “infection and treatment” method using live bacteria, infection with *in vitro* attenuated bacteria [14], immunization with inactivated *in vitro* grown bacteria [15] and recombinant vaccines [16,17] (Table 1). At the moment, the only commercially available vaccine is based on live Ball3 organisms (which are less virulent than the other strains) derived from the blood of infected sheep, and its use is limited to South Africa [18]. The vaccine is administered intravenously and, following inoculation, body temperature is monitored and antibiotic treatment has to be applied when a rise in temperature occurs, to prevent a more serious course of the disease or animal death. One of the most promising alternatives is the inactivated vaccine; not only because it has proven to protect against homologous and heterologous challenge under controlled conditions but also a complete industrial process is now readily available for the production, purification and formulation of large amounts of *E. ruminantium* at low cost [19,20]. Nevertheless, the problems caused by antigenic strain differences and high diversity shown in restricted areas still remain, hampering the development of a fully effective vaccine [21,22].

Globally, most of the vaccines currently available to overcome TBDs are live (blood-derived or attenuated) vaccines (Table 1). Nevertheless they have many drawbacks such as the requirements of a cold chain, a short shelf life and the potential for the transmission of other pathogens and for reversion to virulence [23]. Because of the shortcomings, there is still the need for additional research on the development of alternative safer, more cost-effective and better defined live or subunit vaccines.

Alternative approaches for TBDs control involve the development of anti-tick vaccines [24]. Control of ticks by vaccination has the advantages of being cost-effective, reducing environmental contamination and preventing the selection of drug-resistant ticks that result from repeated acaricide application. The feasibility of controlling tick infestations through immunization of hosts with selected tick antigens was previously demonstrated using recombinant antigens such as Bm86 [25], Bm95, Fer2 [26], Subolesin [27], and EF1a, UBQ chimeric antigens [28]. In some cases, these vaccines also proved to reduce the transmission of TBPs [27,29]. At the moment, two vaccines based on the Bm86 antigen are commercially available: TickGARD (in Australia) and GAVAC (in Cuba and parts of South America), but they are not fully efficacious [30]. As for vaccines against TBPs, the identification of tick antigens inducing host protective response remains the limiting step in the development of effective vaccines limiting tick infestation in hosts [24,31,32].

From the above, it is clear that the development of new generation vaccine is required. For this, it is imperative to improve knowledge of the complex host–vector–pathogen interactions involved in the pathology, immunopathology and protective immune mechanisms of the disease in order to block critical host–pathogen or vector–pathogen interactions. This review highlights how proteomics (associated to other complementary post-genomic tools, such as transcriptomics and metabolomics) can shape further investigation and discovery towards the understanding of such interactions, constituting valuable tools for enhanced rational design of vaccines against these TBDs.

Table 1 – Principal agents of TBDs in cattle, their vector, distribution, cell tropism and vaccine currently available.
Adapted from [1].

Agent	Disease	Principal vectors ^a	Main distribution	Infected cells				Type of vaccine
				In vivo		In vitro		
				Host	Vector	Host	Vector	
<i>Theileria annulata</i>	Tropical theileriosis	<i>H. detritum</i> , <i>H. anatolicum</i>	Northern Africa, Sudan, Southern Europe, Middle East, Indian subcontinent, and part of China	Macrophages	Tick gut cells and salivary glands	Bovine lymphocytes, erythrocytes	(n.a.)	Live and attenuated vaccine (Rakshavac-T®, commercially available)
<i>Theileria parva</i>	East Cost Fever	<i>R. appendiculatus</i>	Eastern, central and southern Africa	Lymphocytes	Tick gut cells and salivary glands	lymphoblast cell lines	(n.a.)	Infection-and-treatment method
<i>Babesia bovis</i>	Tropical bovine babesiosis, Redwater	<i>R. microplus</i> , <i>R. annulatus</i> , <i>R. bursa</i>	Most tropical and subtropical regions	Erythrocytes	Tick gut cells, salivary glands, ovarian tissues	Erythrocytes	(n.a.)	Recombinant: bovipain-2
<i>Babesia bigemina</i>	Redwater, tick fever	<i>R. microplus</i> , <i>R. decoloratus</i> , <i>R. annulatus</i> , <i>R. bursa</i>	Most tropical and subtropical regions	Erythrocytes	Tick gut cells and salivary glands	Erythrocytes	(n.a.)	Live vaccine
<i>Anaplasma marginale</i>	Malignant anaplasmosis, Gallsickness	Most tick species and mechanical transmission by biting flies; congenital transmission	Tropical and subtropical regions	Erythrocytes	tick gut cells and salivary glands	Erythrocyte, endothelial cells	Tick cell lines (such as ISE6, IDE8)	Live and killed vaccines based on <i>A. marginale</i> or <i>A. centrale</i> ; killed vaccine based on tick cell line culture
<i>Ehrlichia ruminantium</i>	Heartwater	10 African <i>Amblyomma</i> spp.	Sub-saharan Africa, Caribbean islands, Madagascar, Comoros	Endothelial cells, neutrophil	Tick gut cells and salivary glands	Endothelial cells, BHK, CHO cells	tick cell line (such as AVL/CTVM13, IDE8, RAE25)	Live, inactivated, and recombinant vaccine (DNA and/or protein-based)

^a Tick genera: H = *Hyalomma*, R = *Rhipicephalus*; “n.a.” stands for “not available” or unknown.

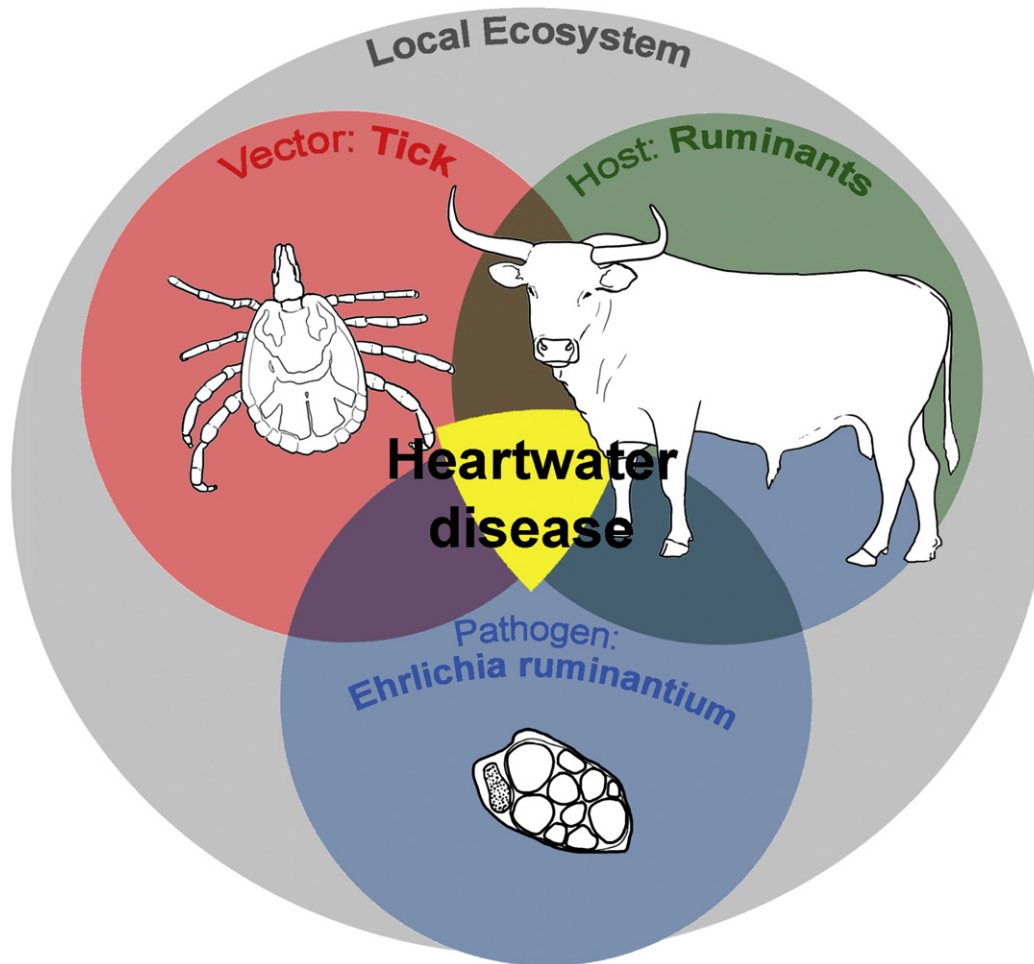


Fig. 1 – The triangle of interactions between tick-borne pathogen, their vector and vertebrate host. The development of new vaccines against tick-borne diseases such as heartwater requires the profound knowledge of the intimate relations between (1) pathogen–tick, (2) pathogen–host and (3) tick–host.

2. Studying host–vector–pathogen interactions: key issues on biological samples

The relationships between TBPs, their tick vectors and diverse vertebrate hosts can be represented by a triangle of parasitic interactions (Fig. 1). In interaction (1) the pathogen interacts with its vector, infecting and replicating within tick cells or extracellular spaces (including those of the gut and salivary glands). In interaction (2), the pathogen interacts with its vertebrate host, infecting and replicating within targeted cells. The third component of the triangle is the interaction between the tick and its host (Fig. 1, interaction 3). Along this review, special emphasis will be given to interactions 1 and 2, as they involve key infection processes related to the adhesion, multiplication and release of the parasite from host and vector cells.

To study pathogen–tick and pathogen–host interactions (interaction 1 and 2 in Fig. 1), biological samples are required, either from *in vivo* or *in vitro* origin (Table 1). One of the restrictions for proteomics research *in vivo* is the limited amount of proteins that can be obtained, namely from the pathogen and tick. In 2010, Villar and co-workers recently showed that Trizol could be

used to extract simultaneous DNA, RNA and proteins from naturally infected and uninfected *Rhipicephalus* spp. ticks to perform Differential In Gel Electrophoresis (DIGE) saturation labeling, but 300 ticks were still required to perform the several assays presented [33].

In vitro cell culture samples have been used, whenever possible, as important alternatives to *in vivo* sampling despite the differential protein expression between the two models [34]. Diverse intracellular pathogens such as *Chlamydomphila pneumoniae* [35], *Plasmodium* spp. [36] or *Leishmania* spp. [37] have been extensively analyzed *in vitro* using either their specific host or vector cells, to mimic key aspects of infectious diseases such as intoxication of host cells by pathogen virulence factors, or bactericidal innate host mechanisms [32].

One of the major constraints when working with obligate intracellular parasites is a large excess of proteins of host or vector origin that can interfere with pathogen protein detection [38]. To overcome this problem, various methods have been used to enrich pathogen proteins. In particular, biochemical fractionation based on differential density and size, harsh detergent treatments dissolving host cells but not pathogen cells, sorting by flow cytometry, or a combination of any of these methods

has been used. Purity and yield of the enriched pathogen samples vary but generally, dozens to hundreds of pathogen proteins could be successfully detected.

Nevertheless, in some cases, both the proteome of the parasite and infected host cells are interesting, as we will discuss afterwards, since infection processes can cause changes in the host cells but cells can also induce changes in the parasite.

3. Proteomic studies on tick-borne pathogens of cattle

Proteomics aims to the large scale analysis of such proteins. It is a powerful tool for the identification of protein and to study their localization, modifications, function and possible interactions or complexes they can form. Unlike “the genome,” there is no single, static proteome in any organism; instead, there are dynamic collections of proteins in different cells and tissues that display moment-to-moment variations in response to several conditions such as stress or infectious processes. Furthermore, proteomics has recently proven its value to find new vaccine candidates against bacterial species such as uropathogenic *Escherichia coli* (UPEC), serogroup B *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Coxiella burnetii*, *Bacillus anthracis*, *Helicobacter pylori*, *Salmonella enteritidis* and many others [39,40].

This progress could be associated to several improvements achieved at all steps of proteomic analysis: sample preparation protocols, peptide/protein separation methods, mass spectrometry (MS) data collection, data analysis and interpretation.

Classical proteomics combines a gel-based analysis to separate proteins and MS for protein identification. Thus, using one- or two-dimensional gel electrophoresis (1-DE or 2-DE), all proteins or ideally a subset of proteins (*e.g.*, surface-exposed) are separated. The resolved proteins can then be excised from the gel and digested into discrete peptides. The proteins present in the original sample are identified by measuring their peptide masses and fragmentation patterns using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS and comparing the experimental data to the predicted masses and fragmentation patterns of known proteins subjected to the same enzymatic digestion. Proteins in the sample that are not identified by MALDI-TOF MS can be further analyzed by electrospray ionization (ESI). Although it is considered a labor-intensive, low throughput technique with poor reproducibility (specially dealing with membrane and small proteins), 2-DE remains the primary method of separating proteins, enabling the detection of intact proteins and distinguishes with high resolution variants that differ in posttranslational modification [32,41]. Moreover, when coupled to immunological assays, 2-DE may also be used to identify B-cell and T-cell antigens within complex protein mixtures. Nowadays, shotgun proteomics is also being used for protein identification. This proteomic approach involves the separation in liquid phase of digested peptides from a complex protein extract, thus avoiding some of the common problems associated with gel separation of hydrophobic proteins or proteins with extreme mass/pI and usually provides markedly better proteome coverage. All these aspects are highly relevant for the proteome characterization of pathogen material from infection models where sample quantity is usually limited. The relative strengths and weaknesses of these two complementary

methodologies (2-DE versus shotgun proteomics) has been previously extensively discussed [37,42–44]. Another proteomic tool that has gain interest in vaccinology is the use of protein microarrays which contain proteins from the parasite printed on glass slides and offer an unbiased screen to detect antigens that react with sera from infected patients [39].

For an accurate protein identification, the knowledge of pathogen and possibly host and vector nucleotide and protein sequence databases are also important [45].

The annotated genome sequences of *T. parva* [5], *T. annulata* [46], *B. bovis* [47], *A. marginale* [48], and *E. ruminantium* are currently available [49,50], and the sequencing of the *B. bigemina* genome is nearing completion (http://www.sanger.ac.uk/Projects/B_bigemina/) [51]. Bovine genome is also currently available [52,53] and despite the rapid advances in molecular acarology, information on tick protein sequences is extremely restricted [54,55]. Indeed genomic information on ticks is so far available for two tick species: *Ixodes scapularis*, *Rhipicephalus (Boophilus) microplus* therefore limiting the application of proteomics. Therefore, there is an urgent need for more genomic information and knowledge on tick proteins expressed in a variety of tissues, life stages and species (and ideally from more than one species).

In the following sections we will present the major outcomes from proteomics studies so far performed for the six TBPs introduced above and also summarized in Table 2.

3.1. *Theileria parva* and *Theileria annulata*

As an example for *Theileria* spp., *T. parva* life cycle is described in Fig. 2, showing the different developmental stages by which the parasite undergoes during its transition in the ruminant host and the tick vector. Proteomic studies are mostly available for host cells.

3.1.1. Proteomics studies in host cells

Proteomic studies on *T. parva* were first published in 1989 [56,57]. Biosynthetically radiolabelled *T. parva* schizonts were purified from bovine lymphoblastoid cells and their proteins were analyzed by 2-DE and autoradiography [56,57]. The studies aimed first to compare *T. parva* stocks from geographically distinct regions (Mariakani and Muguga stocks from Kenya and Uganda stock, from Uganda) and then analyze schizont-protein spot patterns of the same *Theileria* stabilize cultivated in two different infected cell lines. While no significant difference in protein expression pattern was observed between strains, the authors observed a differential protein expression pattern depending on host cells [56]. These results suggested the possibility that selection of phenotypically different parasites could occur *in vivo* or *in vitro*. Sugimoto et al. [57] also analyzed protein and glycoprotein changes induced in bovine lymphoblasts by infection with *T. parva*. The results showed that ten proteins were found in infected cells but not in uninfected cells, and seven of these were detected in preparations of purified schizonts. Four glycoproteins were also detected on the surface of infected cells labeled with [³H]borohydride while a major glycoprotein present on uninfected cells disappeared or was reduced in infected cells. In the 90s, 2-DE and Western blot (WB) analysis were used to identify *T. parva* immunodominant schizont surface antigen [58] and increased knowledge on humoral immune responses to *T. parva* in cattle [59].

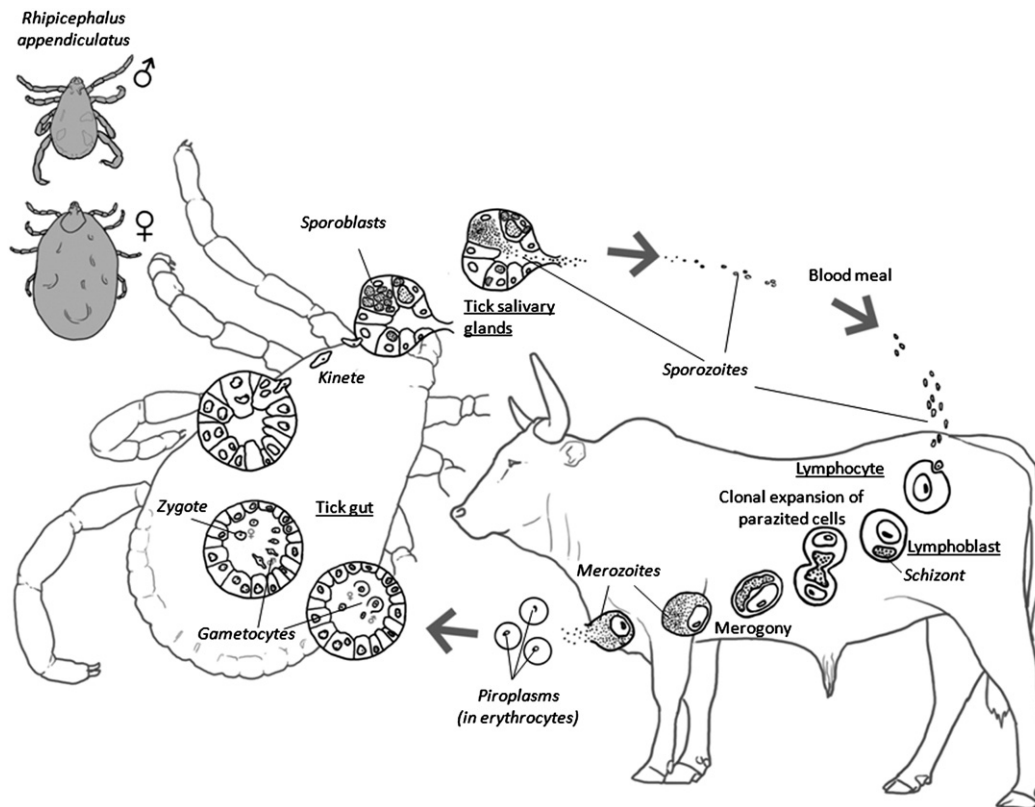


Fig. 2 – *Theileria parva* life cycle. In the animal host, the sporozoites attach to and enter lymphocytes, and develop into forms called schizonts, which infects, multiplies and transforms the white blood cells (lymphocytes or macrophages, for *T. parva* and *T. annulata* respectively) becoming blastoid cells (or lymphoblasts), responsible for disease pathology. Some of the schizonts develop into merozoite forms, which are released from the lymphoblasts into the bloodstream, where they invade erythrocytes and develop into forms called piroplasms. As *Rhipicephalus appendiculatus* ticks feed on animals infected with the parasite, they ingest erythrocytes containing the piroplasms, which are able to infect ticks. Once in the tick gut, the parasites differentiate into male and female gametocytes, which fuse to form zygotes. The zygotes differentiate into kinetes, which move to the salivary gland and enter a particular cell type. The sporozoites are introduced into a mammalian host along with tick saliva when the tick feeds, initiating a new cycle of parasite development. Transmission of *T. parva* is strictly trans-stadial as the parasite is transmitted only by the nymphal and adult stages after acquiring infections as feeding larvae or nymphs. Although *T. annulata* has a similar life cycle, it is exclusively transmitted by *Hyalomma* spp. adult ticks and infects macrophages in the mammalian host.

As previously mentioned, *Theileria* spp. transform host cell (leucocytes) inducing uncontrolled proliferation [60,61]. However, the parasite factors responsible for the inhibition of host cell apoptosis or induction of host cell proliferation are unknown. Recently, Oura and co-workers performed studies on infected host cells using differential RNA display and proteomics to elucidate the interaction between *Theileria* spp. and host cells. The authors suggest that to transform host cells, *Theileria* spp. might modulate the ISGylation system, a key mechanisms associated with resistance of host cells to intracellular infection by pathogens, stimulation of the immune response and terminal differentiation of leukemic cells [61].

3.1.2. Proteomics studies in tick cells

To our knowledge, only one report is currently available for *T. parva* proteins expressed in ticks, being mostly related to the validation of DIGE technology with limited amount of detected proteins in adult infected ticks [62].

3.2. *Babesia bovis* and *Babesia bigemina*

B. bovis and *B. bigemina* are apicomplexan parasites closely related to *Plasmodium* spp. (the causative agents of malaria transmitted by mosquitoes) and to *T. parva*, with similar life cycle in both cattle and ticks. The life cycle of *B. bovis* is presented in Fig. 3; as the developmental cycle in the tick is particularly complex, with the parasites undergoing many changes, these were not detailed in Fig. 3. Interestingly, and contrary to *Theileria* spp., most of the proteomics studies published for *Babesia* spp. have been conducted in infected ticks, with few report available in host cells (Table 2).

3.2.1. Proteomic studies in host cells

One of the first attempts to perform a whole “proteomic” study on the surface antigens of *B. bigemina* infected erythrocytes was performed in 1994 [63]. Using gel electrophoresis and WB analysis, the authors showed that the surface of *B. bigemina* infected

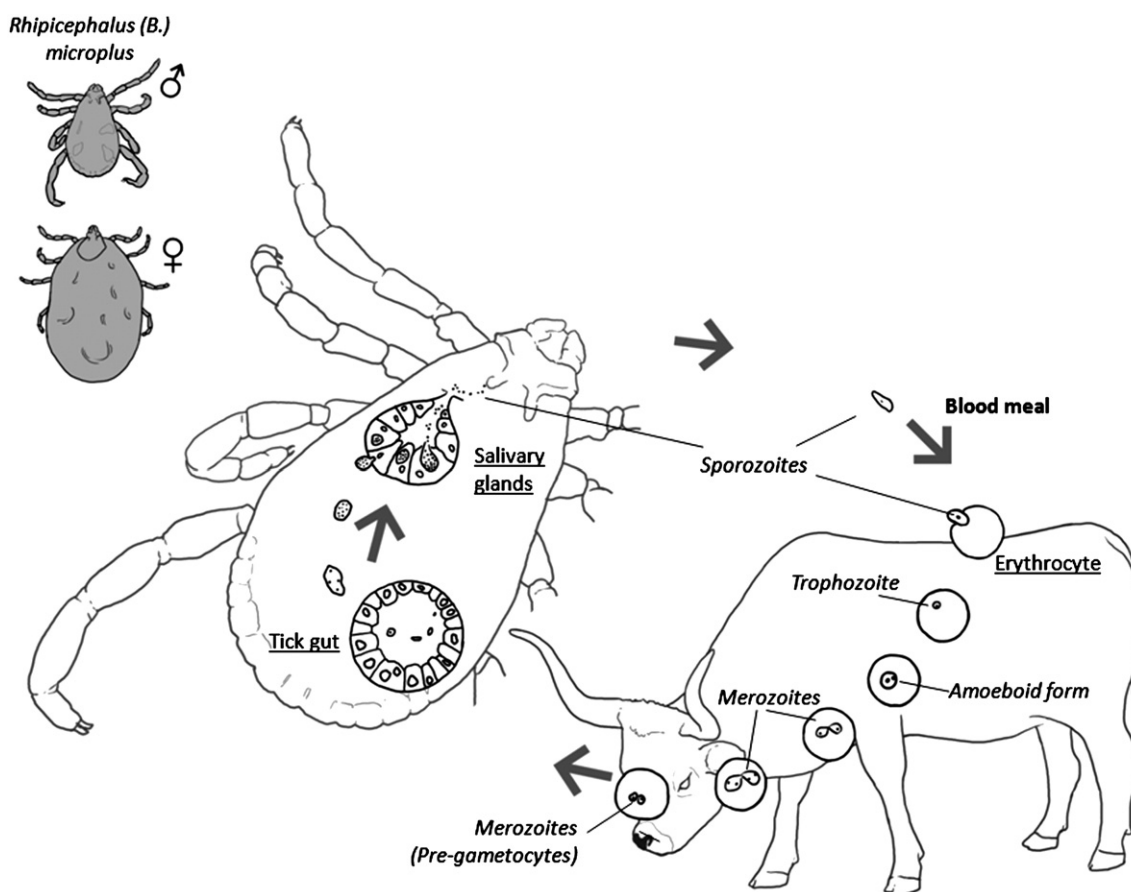


Fig. 3 – *Babesia bovis* life cycle. Infected *Rhipicephalus (Boophilus) microplus* ticks inject the infective sporozoite stage into the mammalian host and the parasites enter erythrocytes where they multiply by binary fission and undergo several changes (trophozoite, amoeboid form) until they become merozoites. After division, merozoites invade other erythrocytes. Ticks become infected by the ingestion of intraerythrocytic parasites. In the gut lumen, the parasites escape from the red cells and invade gut epithelial cells where they undergo massive multiplication. The end result is the production of large parasites called large merozoites, which are released into the hemolymph, which is the tick's 'blood'. Further development outside the intestine occurs in a variety of tissues, the salivary glands and ovaries being especially important for transmission. The merozoites are motile and are able to swim. Some enter the oviduct and invade the developing eggs in the female tick (which are not detailed in this figure). Here, the parasites multiply again and then remain dormant until the eggs hatch and the larval progeny infest a suitable host. After attachment of infected seed ticks, sporozoites in tick salivary glands are injected into the mammalian host at the next blood meal and the *Babesia* is activated and development recommences. The infective forms of *B. bovis* are injected into cattle by larval ticks; those of *B. bigemina* are injected into cattle by nymphal and adult ticks. Sexual development occurs in the tick. *B. bovis* is transmitted transtadially (one tick stage to another stage) but not transovarially.

erythrocytes had parasite-surface exposed epitopes that were conserved among the seven strains examined.

So far, one of the most studied proteins in *B. bovis* is related to a region of synteny with *Theileria* spp. at the p67 and SPAG-1 loci, both highly studied vaccine candidates against ECF and tropical theileriosis, respectively [64]. Freeman and co-workers characterized the *B. bovis*, *bov57* gene in the tick transmissible strain T2Bo by testing for the presence of transcripts in tick and cultured blood stages, but also verifying expression in cultured blood stages using monoclonal antibodies raised against the recombinant protein [64].

3.2.2. Proteomic studies in tick cells

Interestingly, most of the studies on *Babesia* spp. are from tick cell models [65–67]. In 2007, proteomics studies were performed using both gel electrophoresis followed by MS or/and capillary-

HPLC-electrospray tandem mass spectrometry (HPLC-ESI-MS/MS) to investigate differences in expression of soluble and membrane proteins from ovaries of adult female *Rhipicephalus (Boophilus) microplus* ticks — infected or not with *B. bovis* [66]. In 2008, Rachinsky and co-workers performed a similar study in the midgut tissue of uninfected and *B. bovis*-infected *R. microplus* [65]. Among the identified *Babesia*-affected tick midgut proteins six proteins are implicated in signaling processes, including three Ca^{2+} -binding proteins, a guanine nucleotide-binding protein, a protein with signal peptide activity and a translocon-associated receptor protein. Up-regulation of five metabolic enzymes indicated parasite-induced changes in electron and proton transport, protein processing and retinoic acid metabolism. Among the down-regulated proteins were a molecular chaperone, a cytoskeletal protein and a multifunctional protein of the prohibitin family. Identification of these proteins may provide new insights

Table 2 – Transcriptomic and proteomic studies available for the tick-borne pathogens *T. parva*, *T. annulata*, *B. bovis*, *B. bigemina*, *A. marginale* and *E. ruminantium*.

Species	Methodology	Tools	Biological sample	Aim (reference)
<i>Theileria parva</i>	Transcriptomics	Massively Parallel Signature Sequencing (MPSS) Microarray and q-RTPCR cDNA library and EST sequence analysis	Parasite schizont stage <i>T. parva</i> -transformed cell lines Infected tick salivary glands	Whole transcriptome analysis [82] Analysis of subtelomeric variable secreted proteins (SVSP) genes [84] Whole transcriptome analysis [83]
	Proteomics	2D gels, MS, WB	Parasite schizont stage; Infected host cells	Parasite phenotypic characterization [56]; Protein changes in bovine lymphoblastoid cells [56,57]; analysis of surface antigen [58]; characterization of humoral immune responses to <i>T. parva</i> [59].
<i>Theileria annulata</i>	Transcriptomics	Differential mRNA display cDNA microarray (bovine) and RT-PCR	Infected host versus Parasite merozoite stage Infected macrophage cell	Expression analysis during differentiation of <i>T. annulata</i> [87] Effect of two breeds of cattle (resistant and susceptible to disease) [85,86].
	Proteomics	2DE, DIGE	Infected host and tick cells	Modulation of ISGylation system expression [61]; detection in infected tick [33]
<i>Babesia bovis</i>	Transcriptomics	cDNA libraries	<i>In vitro/in vivo</i> host cells; infected tick larvae and egg. Infect host cell	Differential transcription of rRNA [90]; study on antigenic variation [88] Development of an expression oligonucleotide microarray [89]; assess peptidases expression in attenuated versus virulent strain [91]
	Proteomics	Oligonucleotide microarray and q-PCR 2D, LC MS/MS, Western blot	Tick cells	Proteomic profiling of uninfected vs infected tick tissues [65–67]
	Transcriptomics	RT-PCR, Northern blots, and q-RT PCR	Parasite merozoite stage; Parasite sexual and kinetes stages.	Transcription of the rhostry associate protein 1 (rap-1) genes [92]; hsp-20 and rap-1a expression [93]
<i>Babesia bigemina</i>	Proteomics	Immunochemical labeling	Infected erythrocyte	Surface antigens containing epitopes conserved among strains [63]
<i>Anaplasma marginale</i>	Transcriptomics	Microarray and qRT-PCR	Infected Ticks and host cells	Transcription of strain-specific genes at the time of transmission [98]; gene expression in different tick and host cells [71,94–96,98]
	Proteomics	2D and MS	Infected ticks and host cells	Surface proteome on host and/or tick cells [48,68,69,72]
<i>Ehrlichia ruminantium</i>	Transcriptomics	Microarray, q-RTPCR	EB/RB (host and tick infected cells)	SCOTS development [103]; global gene expression profiling of <i>E. ruminantium</i> at different stages of development [104]; map-1 cluster expression [37,99,101,102]
	Proteomics	2D gel and MS	EB/RB (host and tick infected cells)	MAP1-family protein expression [76,77]; first partial elementary proteome in host cells [77].

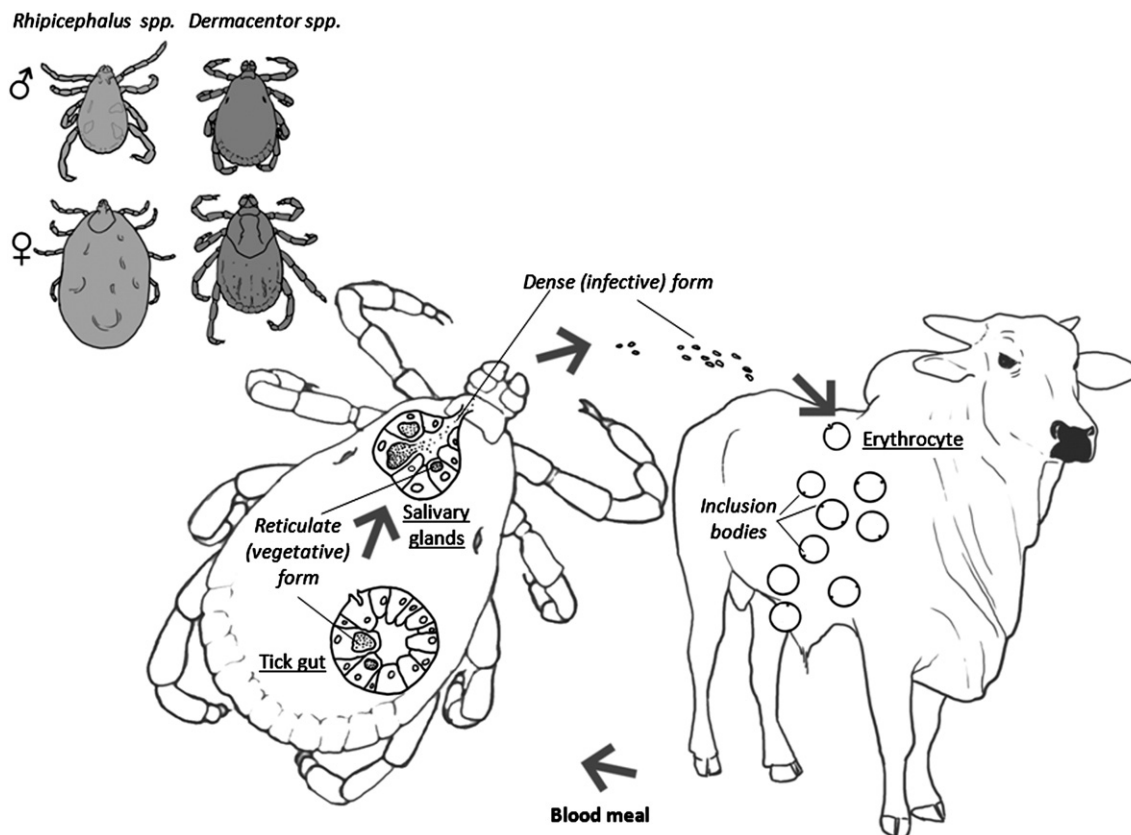


Fig. 4 – *Anaplasma marginale* life cycle. Infected erythrocytes are ingested by ticks (*Dermacentor* spp., *Rhipicephalus* spp.) with the blood meal providing the source of *A. marginale* infection for tick gut cells. After development of *A. marginale* in tick gut cells, many other tick tissues become infected, including the salivary glands, from where the rickettsiae are transmitted to vertebrates during feeding. Two forms of *A. marginale*, reticulated and dense forms, are found in infected tick cells. The reticulated (vegetative) form divides by binary fission forming large colonies that may contain hundreds of organisms; it then changes into the dense infective form and can survive outside the host cells. Cattle become infected with *A. marginale* when the dense form is transmitted during tick feeding via the salivary glands. Tick transmission can occur from stage to stage (transtadial) or within a stage (intrastadial), while transovarial transmission from one tick generation to the next does not appear to occur. Larvae, nymphs and adult ticks can all transmit *A. marginale*. Procedures such as dehorning, castration, vaccination and collection of blood samples may also spread the disease within a herd by indirect contact (use of same surgical instruments without decontamination procedures) [9]. Biting flies can also spread anaplasmosis [9].

into the molecular interactions between *B. bovis* and its tick vector *Rhipicephalus* (*Boophilus*) *microplus*, and could lead to identification of anti-tick and transmission-blocking vaccine candidates [65]. In 2010, additional studies were performed using midgut from partially fed adult female cattle ticks which were analyzed using a combination of 2-DE and gel-free liquid chromatography (LC)-MS/MS [67]. The authors found novel proteins such as clathrin-adaptor protein (involved in the assembly of clathrin-coated vesicles) and membrane-associated trafficking proteins such as Syntaxin 6 and Surfeit 4.

3.3. *Anaplasma marginale*

In stained blood from infected cattle, *A. marginale* organisms viewed under an optical microscope are seen as black, irregular shaped dots, usually at the edge of infected red blood cells (as exemplified in Fig. 4). Nevertheless, little is known about the development cycle of *A. marginale* in ticks, although, it has been observed that ticks do become infected when they ingest

infected red cells and they can retain the infection for at least several weeks.

3.3.1. Proteomic studies in host cells

Proteomic studies on *A. marginale* have been mainly targeted to identify surface proteins as these proteins are known to induce protective immune response in cattle and to understand the transition from the host to the tick vector. Outer membrane protein preparations have been demonstrated to elicit protective immunity for *A. marginale* [48]. In 2005, Lopez and co-workers identified some immunogenic proteins from the purified outer membrane protein complex using 2-DE to separate the proteins, performing WB analysis with sera from cattle immunized with the protective *A. marginale* outer membrane proteins. The immunoreactive proteins were then excised and subjected to LC-MS for definitive identification by mapping the annotated genome [68]. This approach facilitated the identification of 24 immunodominant proteins including the previously characterized major outer membrane proteins Msp2, Msp3, and Msp5 but also proteins

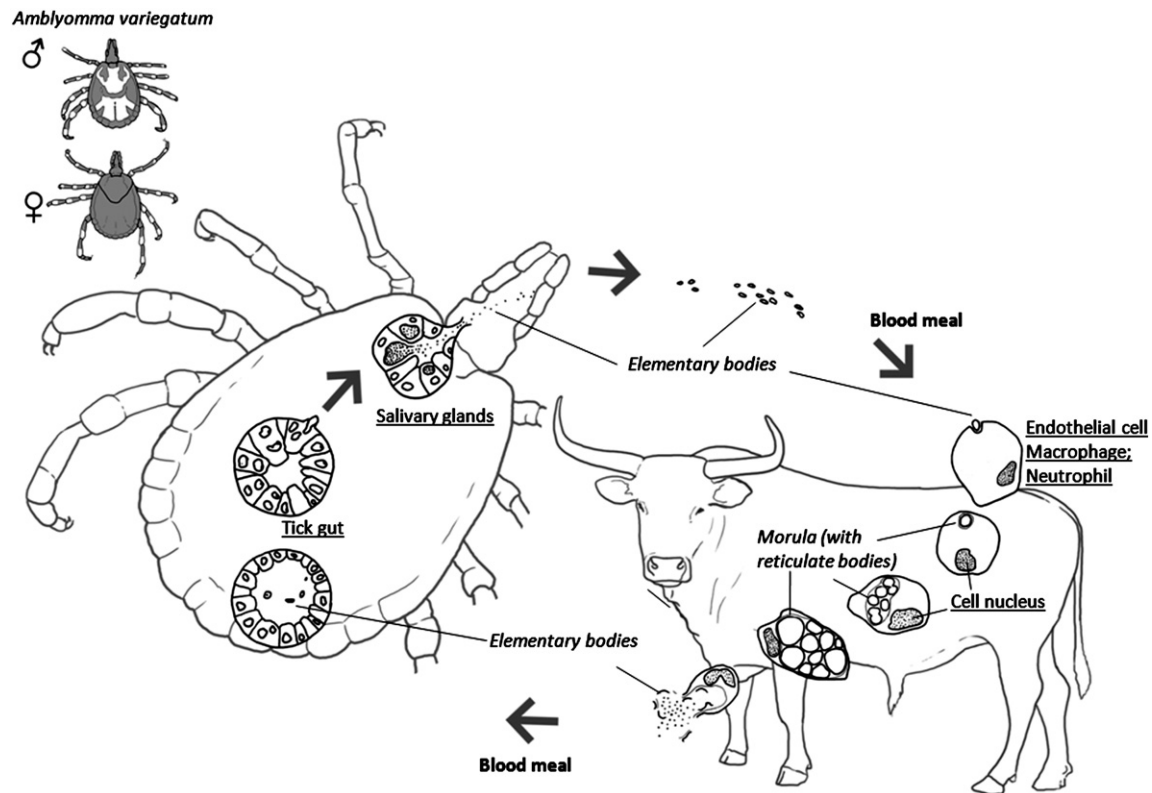


Fig. 5 – *Ehrlichia ruminantium* life cycle. *E. ruminantium* organisms initially develop in the gut epithelial cells of ticks and subsequently invade and develop in the salivary gland cells of the vector (mainly *Amblyomma variegatum*). The vertebrate host is infected via salivary glands of the tick during blood meal. In the vertebrate host, *E. ruminantium* proliferate in vascular endothelial cells, neutrophils and macrophages, presenting a biphasic developmental cycle with two morphologically distinct forms, the elementary bodies (EBs) and the reticulate body (RBs). Organisms enter cells as EBs through a process resembling phagocytosis and divide by binary fission within intracytoplasmic vacuoles resulting in large colonies of RBs (morulae). *Amblyomma* ticks are three-host ticks and become infected during the larval and nymphal stages when they feed on infected hosts. Nymphal and adult ticks transmit *E. ruminantium* to susceptible hosts without losing their infective condition. Still the development cycle of the organism in the tick and the infectivity of successive stages of the tick are poorly understood. Intrastadial transmission has been demonstrated, but transovarial transmission has only been reported once under laboratory conditions and probably does not occur in nature [126].

from the type IV secretion system (TFSS), proteins VirB9, VirB10, and conjugal transfer protein (CTP). A complementary study was published later in 2008 to evaluate the effect of cross-linked surface protein complexes to induce protection against high-level bacteremia and anemia upon *A. marginale* challenge of cattle [69]. For this, intact *A. marginale* were isolated and treated with a membrane impermeable cross-linking reagent, which resulted in covalent linkage of a group of surface exposed outer membrane proteins. After purification, the protein complex was then used to immunize cattle, and was analyzed by MS [69]. This protein complex included only a subset of the complex outer membrane immunogen, comprising Omp1, Omp7–9, Msp1a, Msp2, Msp3, Msp4, OpAG2 and Am779. These results indicate that a surface protein subset of the outer membrane cross-linked using special molecules is capable of inducing protective immunity and serves to direct vaccine development. In 2007, Lopez et al. [70] demonstrate that in *A. marginale* outer membrane-vaccinated cattle, VirB9, VirB10, and CTP are recognized by serum immunoglobulin G2 (IgG2) and stimulate memory T-lymphocyte proliferation and gamma interferon (IFN- γ) secretion

providing a strong support for the use of these protein as potential vaccine candidate against *A. marginale*.

3.3.2. Proteomic studies in tick cells

Several studies were performed to analyze differentially regulated proteins between *A. marginale*-infected and uninfected tick cells.

In 2007, the authors used DIGE and MALDI-TOF/TOF MS to compare the protein expression pattern of *A. marginale*-infected and uninfected IDE8 tick cells. Of the 17 differentially expressed proteins, 7 were from *A. marginale*, 7 could not be identified, and 3 were of tick origin and had homology to sequence databases [71]. Only one protein, homologous to translation elongation factor 1 γ , was up-regulated in infected cells. This study was also complemented by gene expression analysis, as discussed afterwards.

In 2008, Noh and co-workers also analyzed the surface proteome of *A. marginale* isolated from ISE6 tick cells [69], showing that it is less complex than in host cells. Indeed, although, fifteen proteins were identified on *A. marginale* from bovine erythrocytes,

only five were found to be expressed in *A. marginale* from the tick cells. Four proteins, Msp2, Msp3, Msp4, and Am854, a peptidoglycan-associated protein, were expressed in common, Am778 being the only protein to be expressed on the *A. marginale* isolated from tick cells. The authors therefore suggested that remodeling of the surface proteome accompanies the transition between mammalian and arthropod hosts and identify novel targets for blocking transmission [69].

More recently, Ramabu et al. [72] also analyzed *A. marginale* proteins specifically upregulated in ticks in contrast to the mammalian host. Proteins from *A. marginale*-infected and uninfected tick cells and infected bovine erythrocytes were separated using 2-DE [72]. Spots, which were unique to infected tick cells, were submitted for MS analysis. From those spots, 15 *A. marginale* proteins, all annotated as hypothetical were identified, including the previously identified Am778 as well as Am638, an ankyrin-repeat containing protein. Additionally, many metabolic and housekeeping genes were identified, including, but not limited to *dnaK*, *groEL*, and *rpA*. Two proteins from the TFSS were also identified, VirB10 and VirB11. Confirmation of the up regulation in tick cells of a subset of these proteins, including Am470, Am410, and Am829 as compared to bovine erythrocytes was done using WB and densitometry.

3.4. *Ehrlichia ruminantium*

E. ruminantium life cycle has been well described in host endothelial cells [73,74]. Briefly, it has biphasic life cycle with two morphologically distinct forms: the elementary bodies (EB), the extracellular infectious forms of the bacterium and the reticulate bodies (RB), which are intracellular, non-infectious and metabolically active. RBs re-condense back into EBs towards the end of the cycle and are then released from the host cell (Fig. 5). It is known that *Amblyomma* spp. can become infected during the larval and nymphal stages when they feed on infected hosts [75] but *E. ruminantium* development cycle in the tick is poorly understood.

3.4.1. Proteomic studies in host and tick cells

Though several reports are available on gene expression studies in *E. ruminantium* (as discussed below), to our knowledge, only two proteomic studies on *E. ruminantium* are currently available.

In 2008, Postigo and co-workers used 2-DE/MS to analyze the differential expression of the immunodominant *E. ruminantium* MAP1 family proteins in infected bovine endothelial and tick cell cultures. The authors showed that the major proteins detected were MAP1 in *E. ruminantium*-infected endothelial cells, and MAP1-1 in infected tick cells. The authors therefore suggest that this difference in protein expression might indicate that the *map1* multigene family is involved in the adaptation of *E. ruminantium* to the mammalian host and vector tick [76].

In 2011, our group used 2-DE and MALDI-TOF-TOF MS to establish the first 2-DE proteome map of *E. ruminantium* cultivated in endothelial cells (Fig. 6) [77]. Interestingly, amongst the sixty-four spots identified, only four proteins belonging to the MAP1-family were identified (Fig. 6); the other proteins detected were mainly related to energy, amino acid and general metabolism (26%), protein turnover, chaperones and survival (21%) and to information processes (14%) or classified as hypothetical proteins (23%). Additionally, 25% of the detected

proteins were found to be isoforms suggesting that post-translational modifications might be important in EB for the regulation of cellular processes such as host cell recognition, signaling, metabolism and in determining antigenicity as previously observed in other *Rickettsiales*. Additional studies on MAP1-family protein using immunochemical labeling also revealed that these proteins are differentially expressed along the bacterium life cycle, presenting different structural organization. Interestingly, when infectious EBs are released from host cells, MAP1 appears to be organized in SDS and heat-resistant dimers and trimers stabilized by disulfide bridges [77].

4. Complementing proteomics with transcriptomics analyses

Despite significant advances in protein separation and detection techniques and in the accuracy and sensitivity of MS, proteomics

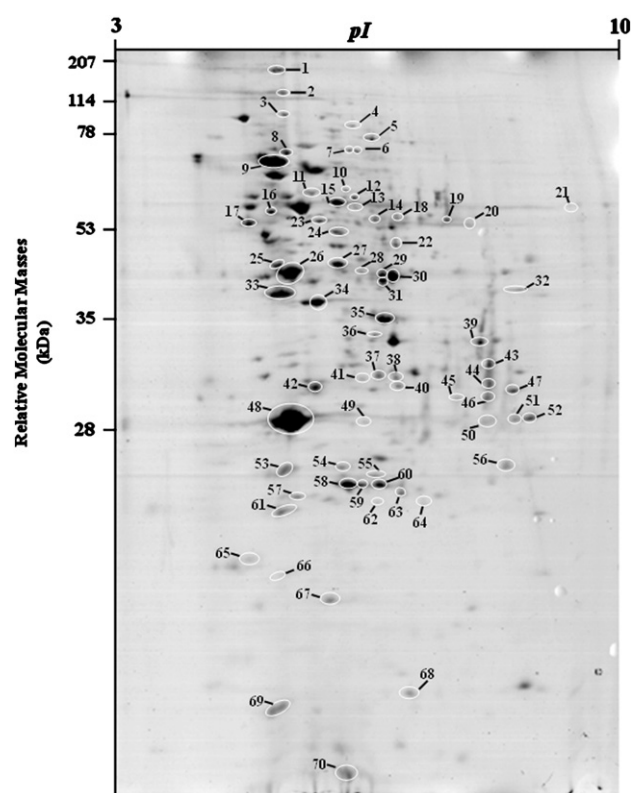


Fig. 6 – Two-dimensional electrophoretic map of infectious *E. ruminantium* Gardel elementary bodies (EBs) proteins expressed at 120 hpi. The crude extract of EB were separated using a pH 3–10 IPG strip in the first dimension, followed by a 12.5% SDS-PAGE in the second dimension. The 2-D gels were stained using Colloidal Coomassie Blue. A representative gel of four different EB batches is shown. Position of molecular masses markers (in kDa) are indicated on the left side. Amongst the sixty-four spots detected, we could identify only four proteins from the MAP1-family: MAP 1–6 (spot 21), MAP 1+1 (spots 43, 44, 45, 46, 62), MAP 1–14 (spot 37) and MAP1 (spots 48, 53), the most abundant spots in the 2D gel image taken from reference [77] published in Veterinary Microbiology.

still presents bottlenecks with the critical disadvantage that, unlike DNA, proteins cannot be amplified to increase the sensitivity of detection techniques. To cope with the difficulties associated to global analysis of protein expression, some researchers still rely on mRNA expression levels as an indicator of the presence of active proteins and use microarrays, real time PCR (qPCR) and reverse transcriptase PCR (RT-PCR) for mRNA detection. Nevertheless, the relationship between mRNA and protein is a complex one and several studies have revealed a relatively weak correlation between mRNA expression levels and protein abundance, with many genes being uniquely detected either by transcriptome or proteome [45,78,79]. This could be due to at least three reasons: i) there are many complicated and varied post-transcriptional mechanisms involved in turning mRNA into protein that are not yet sufficiently well defined to be able to compute protein concentrations from mRNA; ii) proteins may differ substantially in their *in vivo* half lives; and/or iii) there is a significant amount of error and noise in both protein and mRNA experiments that limit the ability to get an accurate picture [80].

As for some TBPs reports on transcriptomics analysis are found in higher number than proteomics ones, we found of interest to also include them in this Review, as complementary information.

4.1. *Theileria* spp.

An interesting review on the *T. parva* available genome information was published in 2000 [81]. The authors briefly present the by then known characteristics of *T. parva* genome and previewed significant insights on vaccine development and disease control through the combination of genome annotation, microarray technology and comparative genomics. *T. parva* transcriptome was analyzed using massively parallel signature sequencing (MPSS) [82]. The authors have established that transcripts were widely distributed throughout the genome and that there was a significant concordance between transcription and protein expression for heat shock proteins, particularly over-expressed in the schizont stage. In 2004, Nene and co-workers analyzed the genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *T. parva* [83]. The resulting EST data are particularly relevant since they can further be used for the construction of microarrays to probe vector biology, vector-host and vector-pathogen interactions and to underpin gene identification via proteomics approaches. In 2009, Schmuckli-Maurer et al. [84] analyzed subtelomeric variable secreted proteins (SVSPs) expression in *T. parva*-transformed cell lines established *in vitro* by infection of T or B lymphocytes. The authors defined that SVSP expression was largely influenced by the parasite genotype and not by the host background or cell type. Interestingly host genotype dramatically influences gene expression of cattle subjected to *T. annulata* infection. In fact, a transcriptomics profiling of *Bos indicus* and *Bos taurus* cultured macrophages infected with *T. annulata* revealed several transcripts with differential expression such as: Toll-like receptor 10 and signal-regulatory protein alpha (SIRPA) or MHC class II DQa, CD9 and prion protein (PRNP) [85,86]. These results clearly demonstrate the importance of the host genetic background in underpinning host response and in a successful disease control.

To identify both host and parasite genes that show altered expression during differentiation of *T. annulata* from the

macroschizont to the merozoite stage of the life cycle, the RNA profiles of two *T. annulata*-infected clonal cell lines (D7 and D7B12) with the same genetic background have been compared by RNA display [87]. After cultivation of D7 and D7B12 macroschizont-infected cells either at 37 °C or at 41 °C (to induce differentiation to the merozoite in the cloned cell line D7), RNA was extracted. From the experiment at 37 °C, 9 cDNA fragments showed altered levels between D7 and D7B12 cell lines, 8 being from host origin, while one was parasite derived. At 41 °C, 6 transcripts showed to be differentially expressed between D7B12 cells and differentiating D7 during a differentiation time course, 1 being of host origin and 5 of parasite origin. Globally, the authors identified a low number of parasite genes involved in *T. annulata* differentiation from macroschizont to the merozoite, but they still recommend the use of RNA display with additional methodologies for further isolation of subsets of differentially expressed genes to provide further insights on this complex differentiation process.

4.2. *Babesia* spp.

In *B. bovis*, Al-Khedery and Allred [88] have conducted a study on the heterodimeric variant erythrocyte surface antigen 1 (*vesa1*) gene which is known to be involved in the virulence, persistence and antigenic variation of the parasite. The authors described the location of the genes and have provided evidence that variation of two transcriptionally active genes occurs through a mechanism of segmental gene conversion involving sequence donor genes of similar organization. The interest of this study to transcriptomics lay in the fact that it combined information from genomics, through the relevant *ves1α* gene and relate it to VESA1 protein and particularly with a key feature of TBPs which is antigenic variation, one of the constraints in vaccine development. In 2007, an expression oligonucleotide microarray was developed showing its potential to analyze *B. bovis* gene expression and hereby complete *B. bovis* infected erythrocyte expression profile [89]. A transcriptional analysis of rRNA gene unit expression in *B. bovis* was published recently [90]. In this study, the authors determine differential transcription of rRNA, depending on both the environment and the life stage of the parasite. Finally, in 2011, Mesplet and colleagues [91] used microarray analysis to the expression of Bovipain-2, a protease released into the host erythrocyte and an important virulence determinant that was found to show differential expression between virulent and attenuated strains.

The *rap-1* gene family occurs in all babesial species examined, rho-try proteins being considered as prime candidates for the development of improved vaccines against bovine babesiosis. In 2003, Suarez and co-workers studied the organization of the complete *rap-1* locus, identifying the *rap-1b* and *rap-1c* genes and examined the expression of all *rap-1* genes in the locus. To test whether all *rap-1* genes in the locus were co-transcribed in merozoites, they used RT-PCR, Northern blots, and quantitative real-time PCR; the results showed that *rap-1a* genes produce the most abundant transcripts of the family, while *rap-1b* transcripts are the least abundant despite the large number of gene copies. Similar patterns of transcription were observed whether merozoites were obtained from *in vitro* cultures or *in vivo* infection. Western blot analysis of merozoites revealed the expected RAP-1a expression but failed to detect

expressed RAP-1b and RAP-1c, indicating that expression of the rap-1 genes is regulated both at the transcriptional and translational levels [92].

In 2008, the expression of *rap-1a* and *hsp20* genes in sexual stages and kinetes of *B. bigemina* was studied using RT-PCR; a complementary expression analysis was carried out using an indirect immunofluorescence test with specific antibodies against HSP-20 and RAP-1a [93]. Globally, the results confirmed the hypothesis that these genes and corresponding proteins are expressed in sexual stages and kinetes, and stress the importance of these proteins in the cellular physiology of tick stages.

4.3. *Anaplasma marginale*

In *A. marginale*, transcriptomics approaches seem to be directed primarily to gene expression of the infected tick vector and infected host cells instead of the pathogen itself. In fact, a functional genomics approach was used to characterize tick genes regulated in response to *A. marginale* infection [71]. According to the authors, four genes of the tick vector, which encode for putative Glutathione S Transferase (GST), salivary selenoprotein M, vATPase, and ubiquitin, are involved in *A. marginale* infection in different tick tissues/organs. These results demonstrated that tick cell gene expression mediates the *A. marginale* developmental cycle and trafficking throughout its vector. More recently, Zivkovic et al. [94] compared tick gene expression in response to two *Anaplasmataceae* species (*A. marginale* and *A. phagocytophilum*) by microarray and qRT-PCR analyses. The results provided evidence of different gene expression responses in tick cells infected with *A. marginale* or *A. phagocytophilum* and the authors concluded that the differences in gene expression and *Anaplasma*–tick interactions reflect differences in the pathogen life cycle at the level of the tick cells. The same research group has also studied differential gene expression at the level of the salivary glands in male *R. microplus* in response to infection with *A. marginale* [95] aiming to expand the knowledge of the molecular mechanisms at the tick–pathogen interface. As a major conclusion, the authors determined that differentially expressed genes encoding for subolesin, putative von Willebrand factor and flagelliform silk protein could play a role in *A. marginale* infection and multiplication in ticks. More recently, Mercado-Curiel et al. analyzed the effect of *A. marginale* infection on the *R. microplus* tick midgut and salivary gland transcriptome during feeding and in response to infection confirming the existence of a massive organ-specific transcriptional response to tick feeding, as numerous *R. microplus* genes were regulated in response to feeding and were differentially regulated between the midgut and salivary gland [96]. Nevertheless, it was found that *A. marginale* exerts a minimal effect on the tick transcriptome.

Previous studies showed that the efficiency of *A. marginale* to develop inside the tick is markedly different for genetically distinct strains [97]. Thus, to confirm this hypothesis Agnes and co-workers analyzed the genes possibly involved in the transmission and infection of two *A. marginale* strains (St. Maries and Israel vaccine) in the tick [99]. The results indicated that although a small difference in gene expression was observed between tick salivary glands infected by St. Maries or Israel vaccine strains during feeding (20 versus 16 genes), the expression levels of certain genes were equal or lower

than those observed in erythrocytes infected by the same strains, suggesting that these genes were not exclusively related to salivary gland colonization.

4.4. *Ehrlichia ruminantium*

In 2002, Bekker and co-workers performed a transcriptional analysis of the MAP-1 (major antigenic protein 1) multigene (*map1-1*, *map1-2* and *map1*) family [99]. In this study, RT-PCR was used to study the transcriptional activity of these genes in different isolates of *E. ruminantium* grown in bovine endothelial cells, in two different tick cell lines, and in *A. variegatum* ticks. The authors concluded that, as for MAP1-1 proteins (mentioned above), the *map1-1* gene transcript was detected in *A. variegatum* ticks, and also in bovine endothelial cells for virulent Gardel and attenuated Senegal. Still, *map1-1* gene was not found in virulent *E. ruminantium* Senegal strain grown in bovine endothelial cells but, being found in different passages of the *in vitro* attenuated Senegal isolate grown in bovine endothelial cells, as well as in the Gardel isolate grown in two tick cell lines. Van Heerden and co-workers have characterized the major outer membrane multigene protein family [100]. Both studies contributed to improve the knowledge on the genetic structure and organization (mapping) of the *map1*-gene family, and the detection of their transcript in different forms of the bacteria either in host or vector cell cultures. In 2005, Bekker and co-workers continued the initial analyses on *map1-1* gene expression and observed that in three *E. ruminantium* strains (Gardel, Senegal and Welgewonden) *map1-1* gene was indeed predominantly detected in infected ticks [101]. The authors finally concluded that the *map1* gene cluster is relatively conserved, but nevertheless subject to recombination. A study on the *map1* multigene family transcription has been carried out *in vivo* in unfed and feeding *A. variegatum* ticks [102]. The results point out that in feeding ticks, *map1-1* transcripts were more abundant in midguts whereas high levels of *map1* transcripts were observed in salivary glands, indicating that the cluster occurs in different tissues before and during transmission, playing therefore a preponderant role in the life cycle of *E. ruminantium*. More recently Emboulé and co-workers optimized the Selective Capture of Transcribed Sequences (SCOTS) methodology to successfully capture *E. ruminantium* mRNAs, avoiding the contaminants of host cell origin and eliminating rRNA which accounts for 80% of total RNA encountered [103]. Overall, the authors conclude that SCOTS has a key importance for *E. ruminantium* transcriptomics analysis (namely for microarrays) and is of potential use in other *Rickettsiales* species, but also in other obligate intracellular bacteria. Pruneau and colleagues [104] have just determined the genome-wide transcriptional profile of *E. ruminantium* replicating inside bovine aortic endothelial cells using cDNA microarrays. Interestingly, over 50 genes were found to have differential expression levels between RBs and EBs. A number of genes involved in metabolism, nutrient exchange and defense mechanisms, including those involved in resistance to oxidative stress, were significantly induced in *E. ruminantium* RBs, a result likely due to the oxidative stress and nutrient starvation occurring in bacteria located in vacuoles. Finally, the authors demonstrate that the transcription factor *dksA*, known to induce virulence in other

microorganisms is over-expressed in the infectious form of *E. ruminantium*.

5. Metabolomics — providing additional insights in pathogen virulence and survival

Metabolomics is the systematic study of all metabolites in a biological cell, tissue, organ or organism, which are the end products of cellular processes. While mRNA gene expression data and proteomic analyses give partial indications of cell modifications, metabolic profiling provides an instant snapshot of its physiology. As with proteomes, metabolomes are dynamic and change in response to nutrition, stress, disease states, and even show diurnal variations. The chemical diversity of metabolome components makes its comprehensive analysis with single analytical technology difficult. The principal technology platforms for metabolomics are NMR spectroscopy and gas chromatography MS (GC–MS) or LC–MS [105].

The metabolome of infectious pathogens has gained a particular interest as researchers have found that the expression of virulence factors might be tightly controlled by their metabolism [106]. For instance, carbon metabolism was found to be crucial for the modulation of the virulence factor PrfA in *Listeria monocytogenes* [107] and certain metabolites involved in iron acquisition are crucial in infections caused by *E. coli* [108]. Additionally, the identification of key metabolites for pathogen development can contribute to develop improved culture medium and, in particular, new axenic medium as previously done for *C. burnetii* [109] and *Tropheryma whipplei* [110]. At the moment, the metabolism of TBPs is poorly known. The published reports are only related to the pyruvate metabolism by *A. marginale* in cell-free culture [111], the amino acid metabolism of *E. ruminantium*-infected endothelial cells [112–115] and the carbohydrate metabolism of *Babesia* spp. in basal salts media [116].

6. System biology: integrating “Omics” data to improve vaccine development

An extraordinary wealth of Omics data generated from genomic, proteomic and metabolomic methodologies provide unprecedented opportunities for biomarker and antigen identification. Accurate understanding of these data and rational prioritization of potential candidates require integrative approaches such as systems biology [117]. To take the full advantages of systems biology methodology, it is necessary to have (i) bioinformatics tools, but also (ii) large data libraries and robust datasets for accurate data mining [118].

Tools such as open reading frame, epitope prediction, and sequence conservation algorithms are commonly used but the implementation of more complex algorithms and software packages is required in order to meld heterogeneous data types coming from the different “Omics” [118]. For instance, modeling software such as BioSignatureDS™ have been used to compare data from both transcriptomic and proteomics approaches to create an overall dynamic of the host–pathogen interaction in order to better understand the host dynamic responses to *Brucella*

melitensis, *Mycobacterium avium paratuberculosis*, *Salmonella enterica* Typhimurium (STM), and a *Salmonella* mutant [117] providing a basis for rational development of vaccines against diseases such as brucellosis and salmonellosis.

Data repositories are also critical to perform an efficient literature mining. As an example is the Vaccine Ontology based approach VO-SciMiner that generated a comprehensive vaccine–gene interaction network, using *Brucella* as a case-study [119]. Information regarding this pathogen as well as many others is also compiled in a web-based vaccine database and analysis system, Vaccine Investigation and Online Information Network (VIOLIN) (<http://www.violinet.org>) [120]. This database compiles both general information regarding pathogens, host range, host protective immunity and vaccine specific information such as analysis of vaccines of commercial use describing type and methods, or vaccine candidates in clinical trials or early stages of development. There are several bioinformatic tools available for data query and analysis, allowing users to search curated data and analyze vaccine-related genes in order to simplify and improve vaccine research. Within VIOLIN, researcher can also use Vaxpresso and Vaxmesh, to search for all possible vaccine-related papers as well as sentences that match specific keywords and ontology-based categories, giving a specific pathogen. Other data repositories include the Gene Expression Omnibus, the Open Proteomics Database, or the Biomolecular Network Database.

Nevertheless, systems biology research is still in its infancy and the maturation of the field will proceed as the many challenges that it faces are addressed and successfully solved. Indeed, despite the increasing availability of data generated by large-scale microarray, proteomic, or sequencing technologies, these must be carefully analyzed and cataloged due to the inherent technical and biological noise in the data set before they can be used for integrative approaches. Furthermore, new predictive computational models that incorporate sufficient molecular details and make effective predictions of pathogen–host or vector–pathogen–host interactions for experimental *in silico* testing need to be developed. This not only involves ample investment cost but also require researchers to be fully engaged in translational research, especially bridging bench and bedside scientists during the development of its own infrastructures and collaborations [121].

Post-genomics tools such as transcriptomics and proteomics and, particularly, metabolomics were only recently being used to study theileriosis, babesiosis, anaplasmosis and heartwater and, of course, the research output is therefore extremely limited for systems biology to be used at the moment. The challenges in using systems biology to TBDS will be related to the integration of “Omics” data from host–vector–pathogen interaction, the different developmental phases of pathogens in both the host and the tick, the parasite attack, host response and parasite counter-response, and the interactions between transmitted pathogen and the vector tick. Another challenge in developing vaccines against these TBDS is related to the wide spectrum of animal hosts in which vaccines must perform effectively; indeed these pathogens can cause diseases in a series of hosts other than cattle such as *Theileria* spp. (infecting African buffalo, Indian water buffalo, waterbuck, and yak) and *E. ruminantium* (which can affect cattle, sheep, goats, and water buffalo).

7. Conclusion

Despite their veterinary importance, no fully effective vaccine is currently available for any of the TBDs theileriosis, babesiosis, anaplasmosis and heartwater.

Until now, vaccines have been mainly developed by following the basic principles established by Louis Pasteur over a century ago: isolation, inactivation, and injection of the causative agent. However, this conventional form of vaccine production is limited in the ability to target pathogens for which no suitable animal model exists and/or bacteria that are antigenically diverse [118].

Nowadays, new breakthroughs in vaccine research are increasingly reliant on novel “Omics” approaches that incorporate high throughput cutting edge technologies such as genomics, proteomics, transcriptomics, and other less known “Omics” such as metabolomics, immunomics, and vaccinomics [122] to deepen our understanding of the key biological processes that lead to protective immunity, observe vaccine responses on a global, systems level, and directly apply the new knowledge gained to the development of next-generation vaccines with improved safety profiles, enhanced efficacy.

Although these strategies have already proven to be useful to develop vaccines against other fastidious pathogens, the use of “Omics” are still in their infancy regarding to TBDs. With the recent availability of TBPs genomes, these high-throughput technologies will now significantly contribute to overcome knowledge gaps on the role of key parasite molecules involved in cell invasion, adhesion, asexual and sexual reproduction, tick transmission and, surely revolutionize the length of time and capacity for discovering potential candidate vaccines (such as proteins involved in protective immune response, tick feeding, or parasite development). The recent development of transient and stable *in vitro* transfection systems for these parasites (such as *B. bovis* [123] and *A. marginale* [124,125]) also pave the way for future exciting developments as they will facilitate the functional analysis of the bacteria genes and will improve our understanding of the biology of and immunological response to these parasites.

By joining all these efforts using an integrative systems biology approach, researchers might then be able to develop a new vaccine against TBDs that would ideally contain antigens from both pathogen and the tick vector to simultaneously avoid the transmission of the pathogens and control tick infestations. Such vaccine would enhance animal productivity, reduce tick control costs, and ultimately improve the standard of living of local populations in susceptible regions.

Acknowledgments

The authors acknowledge the financial support received from the Fundação para a Ciência e Tecnologia PTDC/CVT/114118/2009 research project and PEst-OE/EQB/LA0004/2011 research grant. Author André M. Almeida acknowledges financial support from the Ciência 2007 program, Isabel Marcelino from Post-doc grant SFRH/BPD/45978/2008 all from FCT/MCTES (Lisbon, Portugal). Artwork collaboration by Simão Mateus is greatly appreciated. The authors are also thankful to Dr. Jennifer Pradel for her fruitful comments and suggestions in improving the manuscript.

REFERENCES

- [1] Minjauw B, McLeod A. Tick-borne diseases and poverty. The impact of ticks and tick-borne diseases on the livelihood of small-scale and marginal livestock owners in India and eastern and southern Africa. Edinburgh: Centre for Tropical Veterinary, Medicine, University of Edinburgh, UK; 2003.
- [2] de Castro JJ, James AD, Minjauw B, Di Giulio GU, Permin A, Pegram RG, et al. Long-term studies on the economic impact of ticks on Sanga cattle in Zambia. *Exp Appl Acarol* 1997;21:3–19.
- [3] Babiuk LA. Vaccination: a management tool in veterinary medicine. *Vet J* 2002;164:188–201.
- [4] Lathers CM. Role of veterinary medicine in public health: antibiotic use in food animals and humans and the effect on evolution of antibacterial resistance. *J Clin Pharmacol* 2001;41:595–9.
- [5] Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, et al. Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science* 2005;309:134–7.
- [6] Pipano E, Samish M, Kriegl Y, Yeruham I. Immunization of Friesian cattle against *Theileria annulata* by the infection-treatment method. *Br Vet J* 1981;137:416–20.
- [7] Suarez CE, Noh S. Emerging perspectives in the research of bovine babesiosis and anaplasmosis. *Vet Parasitol* 2011;180:109–25.
- [8] Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, Palmer GH, et al. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc Natl Acad Sci U S A* 2005;102:844–9.
- [9] Kocan KM, de la Fuente J, Guglielmone AA, Melendez RD. Antigens and alternatives for control of *Anaplasma marginale* infection in cattle. *Clin Microbiol Rev* 2003;16:698–712.
- [10] Kocan KM, Halbur T, Blouin EF, Onet V, de la Fuente J, Garcia-Garcia JC, et al. Immunization of cattle with *Anaplasma marginale* derived from tick cell culture. *Vet Parasitol* 2001;102:151–61.
- [11] Palmer GH, Barbet AF, Cantor GH, McGuire TC. Immunization of cattle with the MSP-1 surface protein complex induces protection against a structurally variant *Anaplasma marginale* isolate. *Infect Immun* 1989;57:3666–9.
- [12] Oberem PT, Bezuidenhout JD. The production of heartwater vaccine. *Onderstepoort J Vet Res* 1987;54:485–8.
- [13] Alexander RA. Heartwater. The present state of our knowledge on the disease. Director of Veterinary Services and Animal Industry. South Africa: Union of South Africa; 1931. p. 89–150.
- [14] Zweggarth E, Josemans AI, Van Strijp MF, Lopez-Rebollar L, Van Kleef M, Allsopp BA. An attenuated *Ehrlichia ruminantium* (Welgevonden stock) vaccine protects small ruminants against virulent heartwater challenge. *Vaccine* 2005;23:1695–702.
- [15] Martinez D, Maillard JC, Coisne S, Sheikboudou C, Bensaid A. Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium*. *Vet Immunol Immunopathol* 1994;41:153–63.
- [16] Simbi BH, Bowie MV, McGuire TC, Barbet AF, Mahan SM. Evaluation of *E. ruminantium* genes in DBA/2 mice as potential DNA vaccine candidates for control of heartwater. *Ann N Y Acad Sci* 2006;1078:424–37.
- [17] Sebatiene SI, Pretorius A, Liebenberg J, Steyn H, Van Kleef M. *In vitro* and *in vivo* evaluation of five low molecular weight proteins of *Ehrlichia ruminantium* as potential vaccine components. *Vet Immunol Immunopathol* 2010;137:217–25.
- [18] Allsopp BA. Trends in the control of heartwater. *Onderstepoort J Vet Res* 2009;76:81–8.
- [19] Marcelino I, Sousa MF, Verissimo C, Cunha AE, Carrondo MJ, Alves PM. Process development for the mass production of *Ehrlichia ruminantium*. *Vaccine* 2006;24:1716–25.

- [20] Marcelino I, Vachieri N, Amaral AI, Roldao A, Lefrancois T, Carrondo MJ, et al. Effect of the purification process and the storage conditions on the efficacy of an inactivated vaccine against heartwater. *Vaccine* 2007;25:4903–13.
- [21] Allsopp MT, Allsopp BA. Extensive genetic recombination occurs in the field between different genotypes of *Ehrlichia ruminantium*. *Vet Microbiol* 2007;20:58–65.
- [22] Barbet AF, Byrom B, Mahan SM. Diversity of *Ehrlichia ruminantium* major antigenic protein 1–2 in field isolates and infected sheep. *Infect Immun* 2009;77:2304–10.
- [23] Shkap V, de Vos AJ, Zweygarth E, Jongejan F. Attenuated vaccines for tropical theileriosis, babesiosis and heartwater: the continuing necessity. *Trends Parasitol* 2007;23:420–6.
- [24] de la Fuente J, Kocan KM, Blouin EF. Tick vaccines and the transmission of tick-borne pathogens. *Vet Res Commun* 2007;31(Suppl. 1):85–90.
- [25] Pipano E, Alekceev E, Galker F, Fish L, Samish M, Shkap V. Immunity against *Boophilus annulatus* induced by the Bm86 (Tick-GARD) vaccine. *Exp Appl Acarol* 2003;29:141–9.
- [26] Hajdusek O, Almazan C, Loosova G, Villar M, Canales M, Grubhoffer L, et al. Characterization of ferritin 2 for the control of tick infestations. *Vaccine* 2010;28:2993–8.
- [27] de la Fuente J, Almazan C, Blouin EF, Naranjo V, Kocan KM. Reduction of tick infections with *Anaplasma marginale* and *A. phagocytophilum* by targeting the tick protective antigen subolesin. *Parasitol Res* 2006;100:85–91.
- [28] Almazan C, Moreno-Cantu O, Moreno-Cid JA, Galindo RC, Canales M, Villar M, et al. Control of tick infestations in cattle vaccinated with bacterial membranes containing surface-exposed tick protective antigens. *Vaccine* 2012;30:265–72.
- [29] Merino O, Almazan C, Canales M, Villar M, Moreno-Cid JA, Galindo RC, et al. Targeting the tick protective antigen subolesin reduces vector infestations and pathogen infection by *Anaplasma marginale* and *Babesia bigemina*. *Vaccine* 2011;29:8575–9.
- [30] Andreotti R, Pedrosa MS, Caetano AR, Martins NF. Comparison of predicted binders in *Rhipicephalus (Boophilus) microplus* intestine protein variants Bm86 Campo Grande strain, Bm86 and Bm95. *Rev Bras Parasitol Vet* 2008;17:93–8.
- [31] de la Fuente J, Kocan KM. Strategies for development of vaccines for control of ixodid tick species. *Parasite Immunol* 2006;28:275–83.
- [32] Bumann D. Pathogen proteomes during infection: a basis for infection research and novel control strategies. *J Proteomics* 2010;73:2267–76.
- [33] Villar M, Torina A, Nunez Y, Zivkovic Z, Marina A, Alongi A, et al. Application of highly sensitive saturation labeling to the analysis of differential protein expression in infected ticks from limited samples. *Proteome Sci* 2010, doi:10.1186/1477-5956-8-43.
- [34] Kuntumalla S, Zhang Q, Braisted JC, Fleischmann RD, Peterson SN, Donohue-Rolfe A, et al. *In vivo* versus *in vitro* protein abundance analysis of *Shigella dysenteriae* type 1 reveals changes in the expression of proteins involved in virulence, stress and energy metabolism. *BMC Microbiol* 2011;11:147.
- [35] Molestina RE, Klein JB, Miller RD, Pierce WH, Ramirez JA, Summersgill JT. Proteomic analysis of differentially expressed *Chlamydia pneumoniae* genes during persistent infection of HEp-2 cells. *Infect Immun* 2002;70:2976–81.
- [36] Lasonder E, Janse CJ, van Gemert GJ, Mair GR, Vermunt AM, Douradinha BG, et al. Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog* 2008;4:e1000195.
- [37] Paape D, Lippuner C, Schmid M, Ackermann R, Barrios-Llerena ME, Zimny-Arndt U, et al. Transgenic, fluorescent *Leishmania mexicana* allow direct analysis of the proteome of intracellular amastigotes. *Mol Cell Proteomics* 2008;7:1688–701.
- [38] Li Q, Lostumbo G. Proteomic analyses of a variety of intracellular bacterial species infecting different host cell lines. *Curr Proteomics* 2010;7:222–32.
- [39] Walters MS, Mobley HL. Bacterial proteomics and identification of potential vaccine targets. *Expert Rev Proteomics* 2010;7:181–4.
- [40] Jagusztyń-Krynica EK, Dadlez M, Grabowska A, Roszczenko P. Proteomic technology in the design of new effective antibacterial vaccines. *Expert Rev Proteomics* 2009;6:315–30.
- [41] Rabilloud T, Chevallet M, Luche S, Lelong C. Two-dimensional gel electrophoresis in proteomics: past, present and future. *J Proteomics* 2010;73:2064–77.
- [42] Paape D, Barrios-Llerena ME, Le Bihan T, Mackay L, Aebischer T. Gel free analysis of the proteome of intracellular *Leishmania mexicana*. *Mol Biochem Parasitol* 2010;169:108–14.
- [43] Biron DG, Moura H, Marche L, Hughes AL, Thomas F. Towards a new conceptual approach to “parasitoproteomics”. *Trends Parasitol* 2005;21:162–8.
- [44] Lucitt MB, Price TS, Pizarro A, Wu W, Yocum AK, Seiler C, et al. Analysis of the zebrafish proteome during embryonic development. *Mol Cell Proteomics* 2008;7:981–94.
- [45] Hack CJ. Integrated transcriptome and proteome data: the challenges ahead. *Brief Funct Genomic Proteomic* 2004;3:212–9.
- [46] Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W, et al. Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* 2005;309:131–3.
- [47] Brayton KA, Lau AO, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, et al. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog* 2007;3:1401–13.
- [48] Brayton KA, Palmer GH, Brown WC. Genomic and proteomic approaches to vaccine candidate identification for *Anaplasma marginale*. *Expert Rev Vaccines* 2006;5:95–101.
- [49] Collins NE, Liebenberg J, de Villiers EP, Brayton KA, Louw E, Pretorius A, et al. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc Natl Acad Sci U S A* 2005;102:838–43.
- [50] Frutos R, Viari A, Ferraz C, Morgat A, Eychenie S, Kandassamy Y, et al. Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *J Bacteriol* 2006;188:2533–42.
- [51] Martins TM, do Rosario VE, Domingos A. Identification of papain-like cysteine proteases from the bovine piroplasm *Babesia bigemina* and evolutionary relationship of piroplasms C1 family of cysteine proteases. *Exp Parasitol* 2011;127:184–94.
- [52] Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol* 2009;10:R42.
- [53] Reese JT, Childers CP, Sundaram JP, Dickens CM, Childs KL, Vile DC, et al. Bovine genome database: supporting community annotation and analysis of the *Bos taurus* genome. *BMC Genomics* 2010;11:645.
- [54] Nene V. Tick genomics—coming of age. *Front Biosci* 2009;14:2666–73.
- [55] Ullmann AJ, Lima CM, Guerrero FD, Piesman J, Black WC. Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. *Insect Mol Biol* 2005;14:217–22.
- [56] Sugimoto C, Conrad PA, Mutharia L, Dolan TT, Brown WC, Goddeeris BM, et al. Phenotypic characterization of *Theileria parva* schizonts by two-dimensional gel electrophoresis. *Parasitol Res* 1989;76:1–7.
- [57] Sugimoto C, Mutharia LM, Conrad PA, Dolan TT, Brown WC, Goddeeris BM, et al. Protein changes in bovine lymphoblastoid cells induced by infection with the intracellular parasite *Theileria parva*. *Mol Biochem Parasitol* 1989;37:159–69.

- [58] Sugimoto C, Mutharia LM, Brown WC, Pearson TW, Dolan TT, Conrad PA. Analysis of *Theileria parva* immunodominant schizont surface antigen by two-dimensional polyacrylamide gel electrophoresis and immunoblotting. *Parasitol Res* 1992;78:82–5.
- [59] Kishima M, Dolan TT, Njamunggeh RE, Nkongse CG, Nkongse CG, Spooner PR. Humoral immune responses to *Theileria parva* in cattle as measured by two-dimensional western blotting. *Parasitol Res* 1995;81:334–42.
- [60] Shiels B, Langsley G, Weir W, Pain A, McKellar S, Dobbelaere D. Alteration of host cell phenotype by *Theileria annulata* and *Theileria parva*: mining for manipulators in the parasite genomes. *Int J Parasitol* 2006;36:9–21.
- [61] Oura CA, McKellar S, Swan DG, Okan E, Shiels BR. Infection of bovine cells by the protozoan parasite *Theileria annulata* modulates expression of the ISGylation system. *Cell Microbiol* 2006;8:276–88.
- [62] Villar M, Torina A, Nunez Y, Zivkovic Z, Marina A, Alongi A, et al. Application of highly sensitive saturation labeling to the analysis of differential protein expression in infected ticks from limited samples. *Proteome Sci* 2010;8:43.
- [63] Shompole S, McElwain TF, Jasmer DP, Hines SA, Katende J, Musoke AJ, et al. Identification of *Babesia bigemina* infected erythrocyte surface antigens containing epitopes conserved among strains. *Parasite Immunol* 1994;16:119–27.
- [64] Freeman JM, Kappmeyer LS, Ueti MW, McElwain TF, Baszler TV, Echaide I, et al. A *Babesia bovis* gene syntenic to *Theileria parva* p67 is expressed in blood and tick stage parasites. *Vet Parasitol* 2010;173:211–8.
- [65] Rachinsky A, Guerrero FD, Scoles GA. Proteomic profiling of *Rhipicephalus (Boophilus) microplus* midgut responses to infection with *Babesia bovis*. *Vet Parasitol* 2008;152:294–313.
- [66] Rachinsky A, Guerrero FD, Scoles GA. Differential protein expression in ovaries of uninfected and *Babesia*-infected southern cattle ticks, *Rhipicephalus (Boophilus) microplus*. *Insect Biochem Mol Biol* 2007;37:1291–308.
- [67] Kongsuwan K, Josh P, Zhu Y, Pearson R, Gough J, Colgrave ML. Exploring the midgut proteome of partially fed female cattle tick (*Rhipicephalus (Boophilus) microplus*). *J Insect Physiol* 2010;56:212–26.
- [68] Lopez JE, Siems WF, Palmer GH, Brayton KA, McGuire TC, Norimine J, et al. Identification of novel antigenic proteins in a complex *Anaplasma marginale* outer membrane immunogen by mass spectrometry and genomic mapping. *Infect Immun* 2005;73:8109–18.
- [69] Noh SM, Brayton KA, Brown WC, Norimine J, Munske GR, Davitt CM, et al. Composition of the surface proteome of *Anaplasma marginale* and its role in protective immunity induced by outer membrane immunization. *Infect Immun* 2008;76:2219–26.
- [70] Lopez JE, Palmer GH, Brayton KA, Dark MJ, Leach SE, Brown WC. Immunogenicity of *Anaplasma marginale* type IV secretion system proteins in a protective outer membrane vaccine. *Infect Immun* 2007;75:2333–42.
- [71] de la Fuente J, Blouin EF, Manzano-Roman R, Naranjo V, Almazan C, Perez de la Lastra JM, et al. Functional genomic studies of tick cells in response to infection with the cattle pathogen, *Anaplasma marginale*. *Genomics* 2007;90:712–22.
- [72] Ramabu SS, Ueti MW, Brayton KA, Baszler TV, Palmer GH. Identification of *Anaplasma marginale* proteins specifically upregulated during colonization of the tick vector. *Infect Immun* 2010;78:3047–52.
- [73] Jongejan F, Zandbergen TA, van de Wiel PA, de Groot M, Uilenberg G. The tick-borne rickettsia *Cowdria ruminantium* has a *Chlamydia*-like developmental cycle. *Onderstepoort J Vet Res* 1991;58:227–37.
- [74] Marcelino I, Verissimo C, Sousa MF, Carrondo MJ, Alves PM. Characterization of *Ehrlichia ruminantium* replication and release kinetics in endothelial cell cultures. *Vet Microbiol* 2005;110:87–96.
- [75] Jongejan F, Uilenberg G. Ticks and control methods. *Rev Sci Tech* 1994;13:1201–26.
- [76] Postigo M, Taoufik A, Bell-Sakyi L, Bekker CP, de Vries E, Morrison WI, et al. Host cell-specific protein expression in vitro in *Ehrlichia ruminantium*. *Vet Microbiol* 2008;128:136–47.
- [77] Marcelino I, de Almeida AM, Brito C, Meyer DF, Barreto M, Sheikboudou C, et al. Proteomic analyses of *Ehrlichia ruminantium* highlight differential expression of MAP1-family proteins. *Vet Microbiol* 2011, doi:10.1016/j.vetmic.2011.11.022.
- [78] Xia D, Sanderson SJ, Jones AR, Prieto JH, Yates JR, Bromley E, et al. The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol* 2008;9:R116.
- [79] Cox J, Mann M. Is proteomics the new genomics? *Cell* 2007;130:395–8.
- [80] Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 2003;4:117.
- [81] Nene V, Bishop R, Morzaria S, Gardner MJ, Sugimoto C, ole-MoiYoi OK, et al. *Theileria parva* genomics reveals an atypical apicomplexan genome. *Int J Parasitol* 2000;30:465–74.
- [82] Bishop R, Shah T, Pelle R, Hoyle D, Pearson T, Haines L, et al. Analysis of the transcriptome of the protozoan *Theileria parva* using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. *Nucleic Acids Res* 2005;33:5503–11.
- [83] Nene V, Lee D, Kang'a S, Skilton R, Shah T, de Villiers E, et al. Genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *Theileria parva*. *Insect Biochem Mol Biol* 2004;34:1117–28.
- [84] Schmuckli-Maurer J, Casanova C, Schmed S, Affentranger S, Parvanova I, Kang'a S, et al. Expression analysis of the *Theileria parva* subtelomere-encoded variable secreted protein gene family. *PLoS One* 2009;4:e4839.
- [85] Glass EJ, Jensen K. Resistance and susceptibility to a protozoan parasite of cattle—gene expression differences in macrophages from different breeds of cattle. *Vet Immunol Immunopathol* 2007;120:20–30.
- [86] Jensen K, Paxton E, Waddington D, Talbot R, Darghouth MA, Glass EJ. Differences in the transcriptional responses induced by *Theileria annulata* infection in bovine monocytes derived from resistant and susceptible cattle breeds. *Int J Parasitol* 2008;38:313–25.
- [87] Oura CA, Tait A, Shiels BR. *Theileria annulata*: identification, by differential mRNA display, of modulated host and parasite gene expression in cell lines that are competent or attenuated for differentiation to the merozoite. *Exp Parasitol* 2001;98:10–9.
- [88] Al-Khedery B, Allred DR. Antigenic variation in *Babesia bovis* occurs through segmental gene conversion of the ves multigene family, within a bidirectional locus of active transcription. *Mol Microbiol* 2006;59:402–14.
- [89] Lau AO, Tibbals DL, McElwain TF. *Babesia bovis*: the development of an expression oligonucleotide microarray. *Exp Parasitol* 2007;117:93–8.
- [90] Laughery JM, Lau AO, White SN, Howell JM, Suarez CE. *Babesia bovis*: transcriptional analysis of rRNA gene unit expression. *Exp Parasitol* 2009;123:45–50.
- [91] Mesplet M, Palmer GH, Pedroni MJ, Echaide I, Florin-Christensen M, Schnittger L, et al. Genome-wide analysis of peptidase content and expression in a virulent and attenuated *Babesia bovis* strain pair. *Mol Biochem Parasitol* 2011;179:111–3.
- [92] Suarez CE, Palmer GH, Florin-Christensen M, Hines SA, Hotzel I, McElwain TF. Organization, transcription, and expression of rhoptry associated protein genes in the *Babesia bigemina* rap-1 locus. *Mol Biochem Parasitol* 2003;127:101–12.

- [93] Vichido R, Falcon A, Ramos JA, Alvarez A, Figueroa JV, Norimine J, et al. Expression analysis of heat shock protein 20 and rhoptry-associated protein 1a in sexual stages and kinetes of *Babesia bigemina*. *Ann N Y Acad Sci* 2008;1149:136–40.
- [94] Zivkovic Z, Blouin EF, Manzano-Roman R, Almazan C, Naranjo V, Massung RF, et al. *Anaplasma phagocytophilum* and *Anaplasma marginale* elicit different gene expression responses in cultured tick cells. *Comp Funct Genomics* 2009;705034.
- [95] Zivkovic Z, Esteves E, Almazan C, Daffre S, Nijhof AM, Kocan KM, et al. Differential expression of genes in salivary glands of male *Rhipicephalus (Boophilus) microplus* in response to infection with *Anaplasma marginale*. *BMC Genomics* 2010;11:186.
- [96] Mercado-Curiel RF, Palmer GH, Guerrero FD, Brayton KA. Temporal characterisation of the organ-specific *Rhipicephalus microplus* transcriptional response to *Anaplasma marginale* infection. *Int J Parasitol* 2011;41:851–60.
- [97] Ueti MW, Knowles DP, Davitt CM, Scoles GA, Baszler TV, Palmer GH. Quantitative differences in salivary pathogen load during tick transmission underlie strain-specific variation in transmission efficiency of *Anaplasma marginale*. *Infect Immun* 2009;77:70–5.
- [98] Agnes JT, Herndon D, Ueti MW, Ramabu SS, Evans M, Brayton KA, et al. Association of pathogen strain-specific gene transcription and transmission efficiency phenotype of *Anaplasma marginale*. *Infect Immun* 2010;78:2446–53.
- [99] Bekker CP, Bell-Sakyi L, Paxton EA, Martinez D, Bensaid A, Jongejan F. Transcriptional analysis of the major antigenic protein 1 multigene family of *Cowdria ruminantium*. *Gene* 2002;285:193–201.
- [100] van Heerden H, Collins NE, Brayton KA, Rademeyer C, Allsopp BA. Characterization of a major outer membrane protein multigene family in *Ehrlichia ruminantium*. *Gene* 2004;330:159–68.
- [101] Bekker CP, Postigo M, Taoufik A, Bell-Sakyi L, Ferraz C, Martinez D, et al. Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. *J Bacteriol* 2005;187:4782–91.
- [102] Postigo M, Taoufik A, Bell-Sakyi L, de Vries E, Morrison WI, Jongejan F. Differential transcription of the major antigenic protein 1 multigene family of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks. *Vet Microbiol* 2007;122:298–305.
- [103] Emboule L, Daigle F, Meyer DF, Mari B, Pinarello V, Sheikboudou C, et al. Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol Biol* 2009;10:111.
- [104] Pruneau L, Emboule L, Gely P, Marcelino I, Mari B, Pinarello V, et al. Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol Med Microbiol* 2012;64:66–73.
- [105] Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007;26:51–78.
- [106] Fuchs TM, Eisenreich W, Heesemann J, Goebel W. Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. *FEMS Microbiol Rev* 2012;36:435–62.
- [107] Eisenreich W, Dandekar T, Heesemann J, Goebel W. Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat Rev Microbiol* 2010;8:401–12.
- [108] Henderson JP, Crowley JR, Pinkner JS, Walker JN, Tsukayama P, Stamm WE, et al. Quantitative metabolomics reveals an epigenetic blueprint for iron acquisition in uropathogenic *Escherichia coli*. *PLoS Pathog* 2009;5:e1000305.
- [109] Omsland A, Cockrell DC, Howe D, Fischer ER, Virtaneva K, Sturdevant DE, et al. Host cell-free growth of the Q fever bacterium *Coxiella burnetii*. *Proc Natl Acad Sci U S A* 2009;106:4430–4.
- [110] Renesto P, Crapoulet N, Ogata H, La Scola B, Vestris G, Claverie JM, et al. Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet* 2003;362:447–9.
- [111] McHolland LE, Caldwell DR. Pyruvate metabolism by *Anaplasma marginale* in cell-free culture. *Can J Microbiol* 1999;45:185–9.
- [112] Neitz AW, Yunker CE. Amino acid and protein depletion in medium of cell cultures infected with *Cowdria ruminantium*. *Ann N Y Acad Sci* 1996;791:24–34.
- [113] Josemans AI, Zweggarth E. Amino acid content of cell cultures infected with *Cowdria ruminantium* propagated in a protein-free medium. *Ann N Y Acad Sci* 2002;969:141–6.
- [114] Zweggarth E, Josemans AI. A chemically defined medium for the growth of *Cowdria ruminantium*. *Onderstepoort J Vet Res* 2001;68:37–40.
- [115] Zweggarth E, Josemans AI, Horn E. Serum-free media for the in vitro cultivation of *Cowdria ruminantium*. *Ann N Y Acad Sci* 1998;849:307–12.
- [116] Barry DN. Metabolism of *Babesia* parasites in vitro. Glucose and energy-metabolism and survival of *Babesia rodhaini* in a basal medium with and without adenosine. *Aust J Exp Biol Med Sci* 1982;60:159–65.
- [117] Adams LG, Khare S, Lawhon SD, Rossetti CA, Lewin HA, Lipton MS, et al. Enhancing the role of veterinary vaccines reducing zoonotic diseases of humans: linking systems biology with vaccine development. *Vaccine* 2011;29(41):7197–206.
- [118] Oberg AL, Kennedy RB, Li P, Ovsyannikova IG, Poland GA. Systems biology approaches to new vaccine development. *Curr Opin Immunol* 2011;23:436–43.
- [119] Hur J, Xiang Z, Feldman EL, He Y. Ontology-based *Brucella* vaccine literature indexing and systematic analysis of gene–vaccine association network. *BMC Immunol* 2011;12:49.
- [120] Xiang Z, Todd T, Ku KP, Kovacic BL, Larson CB, Chen F, et al. VIOLIN: vaccine investigation and online information network. *Nucleic Acids Res* 2008;36:D923–8.
- [121] Peng X, Chan EY, Li Y, Diamond DL, Korth MJ, Katze MG. Virus–host interactions: from systems biology to translational research. *Curr Opin Microbiol* 2009;12:432–8.
- [122] Bagnoli F, Baudner B, Mishra RP, Bartolini E, Fiaschi L, Mariotti P, et al. Designing the next generation of vaccines for global public health. *Omics* 2011;15:545–66.
- [123] Suarez CE, McElwain TF. Transient transfection of purified *Babesia bovis* merozoites. *Exp Parasitol* 2008;118:498–504.
- [124] Felsheim RF, Chavez AS, Palmer GH, Crosby L, Barbet AF, Kurti TJ, et al. Transformation of *Anaplasma marginale*. *Vet Parasitol* 2010;167:167–74.
- [125] Noh SM, Ueti MW, Palmer GH, Munderloh UG, Felsheim RF, Brayton KA. Stability and tick transmission phenotype of gfp-transformed *Anaplasma marginale* through a complete in vivo infection cycle. *Appl Environ Microbiol* 2011;77:330–4.
- [126] Bezuidenhout JD, Jacobsz CJ. Proof of transovarial transmission of *Cowdria ruminantium* by *Amblyomma herbraeum*. *Onderstepoort J Vet Res* 1986;53:31–4.

3.2. Article de revue 2: “Omics” approaches in *Rickettsiales*: Recent progress in understanding host-pathogen-vector interactions

L’ordre des *Rickettsiales* est composé de deux familles, la famille des *Rickettsiaceae* dont les membres se développent librement à l’intérieur de la cellule hôte, et la famille des *Anaplasmataceae* dont les membres se développent à l’intérieur d’une vacuole dans la cellule hôte.

La famille des *Anaplasmataceae* inclut 5 genres bien étudiés: *Ehrlichia*, *Anaplasma*, *Aegyptianella*, *Wolbachia* et *Neorickettsia* et deux autres peu étudiés: “*Candidatus Neoehrlichia*” et “*Candidatus Xenohalictis*”. Tous les genres exceptés *Wolbachia* et “*Candidatus Xenohalictis*” sont des pathogènes de mammifères ou d’oiseaux et infectent principalement des cellules d’origine hématopoïétique et/ou des cellules endothéliales.

Chez les *Anaplasmataceae*, les deux espèces les plus étudiées sont *A. phagocytophilum* et *E. chaffeensis*. Elles sont toutes les deux, des pathogènes de l’homme et des animaux.

La famille des *Rickettsiaceae* inclut 2 genres : *Rickettsia* et *Orientia*. Le genre *Rickettsia* est le plus étudié et regroupe des bactéries principalement pathogènes de l’homme. Les deux espèces les plus étudiées sont *R. prowazekii* et *R. conorii* responsables respectivement du typhus exanthématique et de la fièvre boutonneuse méditerranéenne.

Pour cet article de revue, nous avons fait un focus sur les *Rickettsiales* les plus étudiées par les approches « omics ». Il s’agit de *Rickettsia prowazekii*, *Rickettsia conorii*, *Rickettsia felis*, *Anaplasma phagocytophilum* et *Ehrlichia chaffeensis*, qui infectent aussi bien l’homme que les animaux ; *Anaplasma marginale* et *Ehrlichia ruminantium* qui infectent principalement les animaux ; et un endosymbiote des arthropodes et des nematodes, *Wolbachia spp.*

Les analyses de génomique comparative qui consistent à la comparaison des génomes des membres des *Rickettsiales* ou avec d’autres pathogènes intracellulaires obligatoires ou facultatifs, ont permis de mettre en évidence

des déterminants génétiques communs aux *Rickettsiales* ou spécifiques de certains membres. Certains déterminants génétiques sont importants dans les mécanismes d'interactions hôtes-pathogènes-vecteurs tels que le système de sécrétion de type 4 (SST4). Chez *E. chaffeensis* et *A. phagocytophilum* l'importance du SST4 dans la pathogénie a été clairement montrée ([Rikihisa & Lin, 2010](#)). De plus, l'expression des gènes ou la détection des protéines du SST4 ont été mises en évidence pour les *Rickettsiales* étudiées montrant leur expression au cours de l'infection chez l'hôte ou le vecteur.

La comparaison des génomes des *Rickettsiales* a permis de mettre aussi en évidence des clusters de gènes codants pour les protéines de membrane externe, OMP. Leurs nombres varient en fonction des espèces considérées, par exemple, chez *A. phagocytophilum* il en existe plus d'une centaine tandis que chez *E. ruminantium*, elles correspondent à la famille *map* et sont au nombre de 16 ([Dunning Hotopp, et al., 2006](#)). Au niveau du protéome globale des *Rickettsiales*, les OMPs sont les protéines majoritairement détectées. Leur expression a été rapportée comme étant spécifique de l'hôte ou du vecteur, c'est le cas chez *A. phagocytophilum* où la proportion de gènes sur-exprimés codant pour les OMPs était plus importante dans les cellules humaines infectées que dans les cellules de tiques infectées ([Nelson, et al., 2008](#)). Du fait de leur forte expression et de leur polymorphisme caractérisé, les OMPs ont été les premières protéines ciblées pour le développement de stratégies vaccinales.

Des gènes importants pour le développement des pathogènes à l'intérieur de la cellule hôte ont pu être mis en évidence. Il s'agit des gènes codant pour des protéines telles que la superoxyde dismutase ou la thioredoxine permettant une lutte efficace contre le stress oxydatif, ou encore le gène *proP* codant pour un transporteur proline-bétaine impliqués dans la lutte contre le stress osmotique. Les stress oxydatif et osmotique constituent les premiers mécanismes de défense des cellules hôtes contre ces pathogènes ([Dunning Hotopp, et al., 2006](#)). L'expression génique et protéique de ces gènes a pu être mise en évidence pour plusieurs *Rickettsiales* ainsi que pour

notre modèle d'étude *E. ruminantium* ([Marcelino, et al., 2012](#), [Pruneau, et al., 2012](#)).

D'autres gènes codant pour des protéines à motifs eucaryotes sont aussi retrouvés au sein des génomes des *Rickettsiales*. Ces protéines présentant des motifs eucaryotes comme les ankyrines ou les répétitions de tétrapeptides (TRP) sont importantes pour les interactions avec d'autres protéines ou avec de l'ADN ou de l'ARN. Les protéines à motifs ankyrines sont composées de deux hélices alpha séparées par une boucle. Elles peuvent être impliquées dans le cycle cellulaire, l'organisation du cytosquelette, la toxicité ou encore la réponse inflammatoire de l'hôte ([Wakeel, et al., 2010](#)). Par exemple, la protéine AnkA retrouvée chez *A. phagocytophilum* a été identifiée comme étant un effecteur du SST4 et transloquée dans le nucléole de la cellule hôte où elle se fixe sur des promoteurs de gènes de l'hôte, influençant ainsi leur expression ([Al-Khodori, et al., 2010](#)). L'analyse du génome de la souche *E. ruminantium* Gardel, a révélée la présence de gènes codant pour des protéines à motifs ankyrines (CDS_04060 et CDS_06440) ([Frutos, et al., 2006](#)). Les études de génomiques fonctionnelles chez les *Rickettsiales* sont essentiellement réalisées sur la forme infectieuse libre et extracellulaire des bactéries. Ces études permettent de mettre en lumière les gènes impliqués dans l'infection proprement dite et dans les mécanismes d'interaction entre les pathogènes, les hôtes et les vecteurs au moment de l'infection. Cependant, elles ne permettent pas d'identifier les mécanismes mis en jeu par les pathogènes au cours du développement dans les cellules de l'hôte ou du vecteur. De nouvelles recherches sur l'expression des gènes et protéines des *Rickettsiales* au cours du développement seraient intéressantes à la condition d'optimiser les méthodes de purification des échantillons afin d'éliminer les contaminants eucaryotes. D'autre part, la majorité des études est effectuée sur le pathogène mais assez peu sur la réponse de l'hôte au pathogène notamment en protéomique. Pour finir, il existe peu d'études basées sur des comparaisons entre souches virulentes et atténuées. Ce type de comparaison permettrait d'identifier les mécanismes de virulence et

d'atténuation et les gènes impliqués dans l'échappement à la réponse immunitaire de l'hôte.

Omics approaches for understanding *Rickettsiales* pathogenesis

Ludovic Pruneau^{1,2}, Amal Moumène^{1,2}, Damien F. Meyer¹, Isabel Marcelino^{1,3,4}, Thierry Lefrançois⁵ and Nathalie Vachiéry^{1*}

¹ CIRAD, UMR CMAEE, F- 97170 Petit-Bourg, Guadeloupe, France

² Université des Antilles et de la Guyane, Fouillole BP-250, 97157 Pointe-à-Pitre, Guadeloupe, France

³ IBET, Apartado 12, 2780-901 Oeiras, Portugal

⁴ ITQB-UNL, Av. da República, Estação Agronómica Nacional, 2780-157 Oeiras, Portugal

⁵ CIRAD, UMR CMAEE, F-34398 Montpellier, France

*corresponding author: Nathalie Vachiéry e-mail: nathalie.vachery@cirad.fr

E-mail of authors: ludovic.pruneau@cirad.fr, amal.moumene@cirad.fr,

Damien.meyer@cirad.fr, isabel_m31@hotmail.com, Lefrancois.thierry@cirad.fr

Tel: +590-590255442, Fax: +590-590940396

Abstract

The *Rickettsiales* order comprises obligate intracellular bacteria responsible for major human and animal endemic and emerging infectious diseases with important economical and sanitary impacts. These bacteria are transmitted biologically by arthropods. The *Rickettsiales* order comprises the *Anaplasmataceae* and the *Rickettsiaceae* families that have different life cycles inside their host cells. Due to the important impact of *Rickettsiales*, it is essential to better understand their pathogenesis in order to elaborate improved disease control strategies. Here, we review how “Omics” approaches contributed to increase our knowledge on *Rickettsiales* pathogenesis, through a global and integrative strategy to identify genes, transcripts and proteins involved in key pathogen-host-vector biochemical pathways. Comparative genomics aims to identify major gene differences by comparison of genomes of *Rickettsiales* with other intra and extra-cellular bacteria and within a species between strains of different virulent phenotypes. Several genes of *Rickettsiales* coding for Outer Membrane Proteins, Post Translational Modifications, eukaryotic repeated motif proteins, proteins involved to escaping host response (osmotic and oxydative stress) and hypothetical proteins which could play a role in the pathogenesis were identified and their expression was confirmed by functional genomics (mainly transcriptomics and proteomics). We did a gap analysis on “omics” approaches on *Rickettsiales* highlighting that most of the studies have been done on the pathogen mainly on infectious free live forms and not on intracellular replicative forms with a lack of studies on host and vector transcriptome/proteome. Here, we focus our review on 7 *Rickettsiales*, which causes important animal, human and zoonotic diseases: *Ehrlichia ruminantium*, *Anaplasma marginale*, *Ehrlichia chaffeensis*, *Anaplasma phagocytophilum*, *Rickettsia felis*, *Rickettsia prowazekii* and *Rickettsia conorii*. *Wolbachia* spp. which is an endosymbiont of arthropods has been included in this review as a model to compare with other pathogenic *Rickettsiales*.

42 **Keywords (3 to 6):** *Rickettsiales*; omics approaches; genomics; proteomics; transcriptomics

43 Table of contents

44		
45	1) Introduction.....	49
46	2) Comparative genomics to understand <i>Rickettsiales</i> pathogenesis	51
47	2.1) General features.....	51
48	2.2) Genes associated to the pathogenesis of obligatory intracellular <i>Rickettsiales</i>	52
49	2.2.1) Genes involved in the invasion of host cells and escape from host response.....	52
50	2.2.2) Genes involved in the intracellular development	54
51	2.3) Specific genes associated to the pathogenesis of <i>Rickettsiaceae</i> and <i>Anaplasmataceae</i>	
52	54
53	2.3.1) Unique genes coding for outer membrane proteins of <i>Anaplasmataceae</i>	54
54	2.3.2) Unique genes coding for membrane and regulatory proteins of <i>Rickettsiaceae</i>	55
55	2.3.3) Genes involved in host specificity of <i>Rickettsiales</i>	56
56	2.4) Specific genes associated to the pathogenesis by comparison of different strains	57
57	2.4.1) For <i>Ehrlichia chaffeensis</i>	57
58	2.4.2) For <i>Rickettsia prowazekii</i>	57
59	3) Functional genomics	59
60	3.1) General features.....	59
61	3.2) Functional genomics of <i>Rickettsiales</i>	60
62	3.2.1) Strategies and methods to perform transcriptomic and proteomic studies on	
63	<i>Rickettsiales</i>	60
64	3.3) Role of OMPs and post-translational modifications (PTM) during infection.....	62
65	3.3.1) In host cells:.....	62
66	3.3.2) Comparison of OMP gene/protein expression in tick and host cells.....	62
67	3.3.3) Differential PTM of proteins in tick and host cells	63

68	3.4) Identification of Type IV secretion systems (T4SS).....	63
69	3.5) Expression of proteins involved in the subversion of host immune response	64
70	3.6) Importance of hypothetical proteins during infection	65
71	3.7) Identification of pathogen metabolic activities in host and vector.....	66
72	3.8) Functional genomics of host cells in response to infection.....	67
73	3.9) Functional genomics of vector cells in response to infection	68
74	4) Conclusion and future perspectives	70
75	5) Authors' contributions	71
76	6) Acknowledgement and financial support.....	71
77	7) References.....	72
78		
79		

1) Introduction

The *Rickettsiales* order includes small obligate intracellular α -proteobacteria, most closely related to mitochondria [1]. These bacteria are transmitted by blood-sucking arthropods, and are responsible for major endemic and emerging human and animal infectious diseases with important economical and sanitary impacts. The *Rickettsiales* order is divided in two families based on their 16S sequence: the *Anaplasmataceae* and the *Rickettsiaceae* (figure 1). The *Anaplasmataceae* family includes five well-known genera, *Ehrlichia*, *Anaplasma*, *Aegyptianella*, *Wolbachia* and *Neorickettsia* and two less-studied genera, “*Candidatus* Neoehrlichia” and “*Candidatus* Xenohaliois” [2]. These bacteria infect invertebrate hosts that are abundant and ubiquitous in the environment (*ie* ticks, insects, trematodes, nematodes or mollusks). Unlike *Neorickettsia* and *Wolbachia* spp., which can be transmitted through generations of invertebrate hosts by both transovarial and trans-stadial transmission, *Anaplasma* and *Ehrlichia*, that are transmitted respectively by *Ixodes Rhipicephalus* and *Amblyomma* ticks seem to have only trans-stadial transmission [3, 4]. All genera except *Wolbachia* and “*Candidatus* Xenohaliois” are known to infect vertebrates (mammals or birds). The bacteria infect specific host cell types, such as neutrophils, monocytes and macrophages, platelets, erythrocytes or endothelial cells depending on the species. Within the *Anaplasmataceae* family, *A. phagocytophilum* and *E. chaffeensis* are widely studied. *A. phagocytophilum* is responsible for anaplasmosis and infects deers, dogs, cats, horses, ruminants, rodents and humans (named human granulocytic anaplasmosis) (Table 1). *E. chaffeensis* infects deer, dogs and human inducing canine ehrlichiosis and human monocytic ehrlichiosis. The two bacteria infecting only domestic and wild ruminants are *A. marginale* and *E. ruminantium*, causing bovine anaplasmosis and heartwater, respectively. The geographical disease distribution and the main associated vectors are described in table 1. Bovine anaplasmosis has the most widely distribution within tick borne diseases. Acute form

of the disease induces fever, anemia loss of weight and often death. After infection, animals are asymptomatic carriers and constitute a reservoir of transmission of the disease. Heartwater due to *Ehrlichia ruminantium* which is present in Sub-saharan Africa, Caribbean islands and Indian Ocean induce mortality in ruminants and avoids any improvement of productivity in the herds. The economical impact of heartwater associated to mortality and cost treatment (antibiotic and acaricide) is estimated as \$ 47.7 million per year for the Southern community. The *Rickettsiaceae* family comprises two genera: *Rickettsia* and *Orientia* (Figure 1). *Rickettsia* spp. are widely studied as they are mainly human pathogens. The *Rickettsia* genus is divided in four groups: the spotted fever group (SFG) (including *R. rickettsii*, *R. conorii*, *R. africae* and *R. japonica*), the typhus group (with *R. prowazekii* and *R. typhi*), an ancestral group (including *R. bellii* and *R. Canadensis* with no evidence of pathogenicity) and a transitional group (*R. akari*, *R. australis*, and *R. felis*). *R. conorii* (causing Mediterranean spotted fever) and *R. prowazekii* (causing Epidemic typhus) are the two most studied species (Table 1). *Rickettsia* spp. are transmitted by hematophagous arthropods, *i.e.* fleas ticks and lice. Further investigation on a variety of hosts including non-hematophagous insects, amoebae and leeches allowed to identify many novel *Rickettsia* clades, opening a broader view of their evolution [5].

The main difference between *Anaplasmataceae* and *Rickettsiaceae* families resides in the life cycle inside the host. *Anaplasmataceae* have an intra-vacuolar development in the host cell cytoplasm along its intracellular life cycle whereas *Rickettsiaceae* after entrance in the host cell escape from the phagosome and multiply in the cytoplasm before release in the extracellular environment (Figure 2). Different stages of development are defined for *Anaplasmataceae* with DNA reorganization from dense cored cell (infectious form) to reticulate cell (vegetative form), which grows, by binary fission to form morulae and then dense cored cell before releasing.

The understanding of the *Rickettsiales* biology and pathogenesis was strongly hampered by their obligatory intracellular characteristic, resulting in culture constraints and difficulties to study the genetics of these bacteria. During the last 10 years, development of *in vitro* models and significant technical progresses in molecular biology using high throughput methods allowed to generate new knowledge on their genome expression. Here, we review recent advances in the understanding of the *Rickettsiales* pathogenesis, using comparative genomics, transcriptomics and proteomics approaches. We focus our review on the most studied *Rickettsiales* by “Omics” approaches and inducing important animal and zoonotic diseases: *Rickettsia prowazekii*, *Rickettsia conorii*, *Rickettsia felis*, *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* (all infecting humans and animals), *Anaplasma marginale* and *Ehrlichia ruminantium* which strictly infect animals. An endosymbiont of arthropods and nematodes, *Wolbachia* spp. have been included in this review as a model to compare with other pathogenic *Rickettsiales*.

2) Comparative genomics to understand *Rickettsiales* pathogenesis

2.1) General features

According to the genome database Integrated Microbial Genomes [6] the whole genome sequences of 59 strains of *Rickettsiales* are available, including 7 genomes of *Anaplasma* spp., 6 of *Ehrlichia* spp., 11 of *Wolbachia* spp. and 46 of *Rickettsia* spp. *Rickettsiales* have a small genome size (0.86-2.13 Mb) and show a low GC percent less than 36% (except for *Anaplasma* spp. and *Neorickettsia* spp.) with a proportion of coding genome ranging between 64.4% and 86% (Table 2). These genomic characteristics may be related to the intracellular lifestyle of the bacteria which mechanically limit the capacity for gene acquisition by lateral gene transfer [7]. Genome reduction seems also to be associated with increased pathogenicity [1, 8]. Interestingly, *E. ruminantium* displays the lowest coding ratio observed among all bacteria due to unusually long intergenic regions. There is an active process of genome

expansion/contraction targeted at tandem repeats in noncoding regions that is unique to *E. ruminantium* and not observed in the other *Rickettsiales* [9].

Amongst all the sequenced *Rickettsiales*, only *Anaplasma spp.* and *Ehrlichia spp.* showed the same synteny. However, there are three regions conserved among all the *Rickettsiales*, including two operons encoding for ribosomal proteins and one operon encoding for components of the type IV secretion system (T4SS) [10].

The availability of complete genome sequences of different *Rickettsiales* enables comparative genomics to identify major differences and similarities between them. Comparative genomics contributes to a better understanding of bacterial evolution, physiology and pathogenesis.

2.2) Genes associated to the pathogenesis of obligatory intracellular *Rickettsiales*

2.2.1) Genes involved in the invasion of host cells and escape from host response

Most intracellular bacteria, including the *Rickettsiales* contain genes encoding for proteins with eukaryotic-like domains such as ankyrin repeats (ANK) and tetratricopeptide repeats (TPR) [1]. These patterns of eukaryotic origin are encoded within proteins to facilitate protein-protein and protein-nucleic acid interactions, being potentially involved in the pathogenicity of bacteria. For example, *A. phagocytophilum* secretes the AnkA Type IV secreted effector protein into the host cytoplasm. AnkA is secondarily localized in the nucleus where it binds to nuclear proteins to form complexes with AT-rich DNA sequences, some of which are promoter sites, hence influencing host gene transcription [11].

Repetitive sequences without eukaryotic-like domains are also involved in the pathogenicity. *Ehrlichia spp.* genomes have numerous tandem repeats (TRP) [9]. TRP containing proteins in pathogenic bacteria are known to be associated with host-pathogen interactions. For example, TRP47 in *E. chaffeensis* was described as a multi-functional effector that plays an important role in establishing bacterial infection and promoting intracellular survival [12]. TRPs are

179 highly present on *E. ruminantium* genome but their function is so far unknown, deserving
180 further characterization.

181 All *Rickettsiales* genomes contain putative pathogenesis related-genes coding for enzymes
182 counteracting host innate immune response (osmotic and oxidative stress), outer membrane
183 proteins and proteins involved in secretion systems [10, 13]. These proteins are crucial for
184 intracellular bacteria development and for its pathogenesis.

185 All *Rickettsiales* have either a Sec-dependent or a Twin arginine translocation (Tat) (Sec-
186 independent) export pathway for secretion of proteins in the periplasm of the bacterium [10].
187 These pathways can be involved in the transport of hydrolytic enzymes such as
188 phospholipases and proteases, which are important in the pathogenesis. In addition, the Tat
189 pathway might be a characteristic of the parasitic lifestyle of *Rickettsiales* allowing the
190 scavenging of nutrients from the host [14].

191 All *Rickettsiales* have a T4SS, which consists of a multiprotein complex that injects effector
192 proteins into eukaryotic cells. The paradigm of T4SS is that of *Agrobacterium tumefaciens*,
193 which contains 12 *virB/D* genes. Except for *virB1* and *virB5*, all the components of the *A.*
194 *tumefaciens* T4SS are conserved in *Rickettsiales* but can be duplicated and scattered in several
195 gene clusters [15]. The crucial role of T4SS and its effector proteins in the pathogenesis has
196 been shown for *A. phagocytophilum* and *E. chaffeensis* [16].

197 In order to identify specific genes of *Rickettsiales*, the genomes of 19 obligate and facultative
198 intracellular pathogens and endosymbionts were compared [10]. The genomes of *Rickettsiales*
199 studied in this review were compared to other intracellular bacteria such as *Bartonella*
200 *henselae*, *Coxiella burnetii* and *Chlamydia pneumoniae*. Eleven Clusters of Orthologous
201 Genes (COG) were found to be unique to *Rickettsiales* genome. These include a type I
202 secretion system ATPase and the T4SS proteins VirB9 and VirB8. Type I secretion system

allows secretion of toxins, proteases and lipases in the extracellular milieu, and thus can be potentially involved in the pathogenesis of *Rickettsiales*.

2.2.2) Genes involved in the intracellular development

To analyze the specific intracellular characteristics of *Rickettsiales*, Dunning Hotopp and co-workers compared the genomes of some *Anaplasmataceae* and a closely related order of extracellular-living organisms, *Caulobacter crescentus*. The study showed that compared to *C. crescentus*, *A. phagocytophilum*, *E. chaffeensis*, *W. pipientis*, and *N. sennetsu* have a significantly higher percentage of genes involved in the biosynthesis of nucleotides cofactors, vitamins and proteins [10]. Some of these genes encode for ribosomal proteins, tRNA synthetases, RNA modification enzymes and translation factors.

Furthermore, *Anaplasmataceae* species have a significantly lower coding capacity for central intermediary metabolism, transport and regulatory functions. The reduction of genes coding for these functions may reflect their ability to acquire specific compounds from their hosts. The genes involved in biosynthetic pathways are essential for intracellular development and thus are still present despite the genome reduction process.

2.3) Specific genes associated to the pathogenesis of *Rickettsiaceae* and *Anaplasmataceae*

Twelve clusters of conserved hypothetical proteins and one cluster of gene encoding for GCN5-related N-acetyltransferase family (GNAT) enable to distinguish the *Anaplasmataceae* from the *Rickettsiaceae* [10].

2.3.1) Unique genes coding for outer membrane proteins of *Anaplasmataceae*

Outer membrane proteins (OMP) play several important roles in bacteria, allowing them to adapt to different environments and host niches. These roles include biogenesis and integrity of the outer membrane, non specific porin activity, adherence and membrane associated enzymatic activity [17]. Some OMPs are porins that form channels allowing the transport of

227 molecules across lipid bilayer membranes and play a major role in host-interaction. They
 228 contribute to nutrient transport, antimicrobial resistance and response to osmotic stress and are
 229 essential for bacteria [18].

230 OMP in obligate intracellular bacteria seems to be crucial and could facilitate early
 231 interactions with the host and vector cells. Many OMPs in *Anaplasmataceae*, have a Pfam
 232 domain PF01617 and constitute the OMP-1/MSP2/P44 superfamily. *Ehrlichia spp.* and
 233 *Anaplasma spp.* have the largest expansion of this family, whereas *Wolbachia spp.* have the
 234 smallest expansion with only 3 OMPs [10]. Indeed, there are over 100 members in *A.*
 235 *phagocytophilum*, 56 in *A. marginale*, 22 in *E. chaffeensis*, 16 in *E. ruminantium* and 25 in *E.*
 236 *canis*. The presence of such numerous OMPs in the genomes of *Anaplasma spp.* and
 237 *Ehrlichia spp.* is unique and remarkable. Moreover, their organization is conserved among
 238 *Ehrlichia spp.* The largest expansion of this family in *Ehrlichia* and *Anaplasma spp.*, may be
 239 related to the pathogen ability to persist in the the vertebrate reservoir and to be transmitted by
 240 tick. The 3 OMPs identified in *Wolbachia* are Wsp and its two paralogs, WspA and WspB.
 241 Wsp protein elicits an immunological response in vertebrate hosts infected with filarial
 242 nematodes harboring *Wolbachia*, but its function is unknown in arthropods [10]. Further
 243 studies are necessary to better understand the role of OMPs in the *Rickettsiales*.

244 2.3.2) Unique genes coding for membrane and regulatory proteins of Rickettsiaceae

245 For *Rickettsia spp.*, a family of 17 surface cell antigens (*sca*) genes has been identified. The
 246 predicted products of these genes are autotransporter proteins [19]. This family exhibits a high
 247 degree of degeneration, but *sca0*, *sca5*, *sca1*, *sca2* and *sca4* genes are highly conserved
 248 among many SFG *Rickettsiae*. The *rickA* and *sca2* genes (coding for surface cell antigens),
 249 absent in the *Anaplasmataceae*, are present and functional in *Rickettsia spp.* except for *R.*
 250 *prowazekii*, *R. typhi*, *R. peacockii*, and *R. canadensis*. The RickA and Sca2 proteins play a

role in actin-based motility [20, 21]. These results suggest that *sca* genes are essential for different stages in intracellular life [21].

The gene family *spoT* was detected only in *Rickettsia spp.* genomes. The *spoT* gene encodes for global metabolism regulators controlling the concentration of alarmone, which is an intracellular signal molecule, changing the global cellular metabolism in response to a stringent response, such as nutrients restriction. The alarmone system is up-regulated in starvation conditions and permits a better adaptation of *Rickettsia spp.* to the available nutrients, suggesting an active role in adaptation to the host or vector cells [1, 10].

2.3.3) Genes involved in host specificity of Rickettsiales

Comparative genomics of *Rickettsiales* infecting various hosts also permitted to identify coding genes potentially required for specific host infection. Indeed, Dunning Hotopp and co-workers identified hypothetical proteins present only in pathogens infecting strictly animals such as *A. marginale* (AM279) and *E. ruminantium* (ERGA_CDS_01780) and absent in other *Rickettsiales* infecting both human and animal [10]. This observation suggests that AM279 and ERGA_CDS_01780 could play a role in the host specificity. Though this protein is not essential for animal infection, its absence could allow infection in humans. Furthermore, the authors identified 4 orthologous clusters of conserved hypothetical proteins in all the pathogenic *Rickettsiales* and absent in the endosymbionts. It would be then interesting to characterize the functions of these proteins, because they may be essential for pathogenesis and survival in the vertebrate host.

Three orthologous clusters, including one for bacterioferritin and two for conserved hypothetical proteins, are present in *Wolbachia spp.* and *N. sennetsu* but absent in all of the tick-, flea- and louse-borne *Rickettsiales*. These COGs may be linked to the lack of a blood-sucking arthropod in the life cycles of these organisms. One orthologous cluster containing a

class II aldolase/adducing domain protein is absent only in *Ehrlichia spp.* and *Anaplasma spp.*; hereby preventing transovarial transmission in the arthropod vector [10].

2.4) Specific genes associated to the pathogenesis by comparison of different strains

2.4.1) For *Ehrlichia chaffeensis*

To analyze the effect of genomic divergence on *E. chaffeensis* virulence potential, Miura and Rikihisa compared the genomes of three strains with distinct virulence phenotypes (Wakulla being more virulent than Liberty and Arkansas strains) by comparative genome hybridization using a tiled microarrays of the Arkansas strain genome [22]. The results showed that 4,663 and 5,325 positions in the chromosome of the Wakulla and Liberty strain, respectively, were different from those in the chromosome of Arkansas strain, including three common major polymorphic chromosomal regions. The differences observed, were localized on genes coding for cell envelope proteins such as *omp-1 p28* family, for ANK repeat proteins and for hypothetical proteins. These results clearly revealed distinct virulence phenotypes of *E. chaffeensis* strains associated with defined genome sequences [22].

2.4.2) For *Rickettsia prowazekii*

More recently, Bechah et al. compared genome sequences of virulent (Rp22) and avirulent (Madrid E) strains of *R. prowazekii* [23]. They found 81 genes differing between the two strains: 64 with non-synonymous mutations or gaps, 11 complete genes in Rp22 but truncated in Madrid E, four truncated genes in Rp22 but complete in Madrid E, and two truncated genes in both strains, at different position. The number of truncated genes was significantly higher in Madrid E compared to Rp22. These results suggested that mutations inducing truncatures heavily impact on the virulence of *R. prowazekii*. The authors then compared the 81 genes differing between these two strains with those of five additional strains of *R. prowazekii* strains with various degree of virulence: Brein1, Madrid E-USA, Madrid E-M, Evir, and Erus.

300 This allowed the identification of four genotypes associated to the four different phenotypes.
301 Moreover, the authors revealed that the generation of the avirulent mutant seems to be due to
302 the knock-out of the *recO* gene. This gene codes for a protein involved in a protein complex
303 (RecFOR) that plays a major role in DNA reparation. They showed that the restoration of
304 *recO* gene was rapidly followed by the restoration of other genes, which may explain how the
305 virulence may be restored in several steps [23]. More experiments should be required to
306 confirm the exact role of *recO* in the *R. prowazekii* virulence.
307 Similarly to the studies presented above for *E. chaffeensis* and *R. prowazekii*, the differences
308 between virulent and attenuated *E. ruminantium* strains (Senegal, Welgevonden and Gardel),
309 is currently being studied by our group (unpublished data).
310

***Rickettsiales* genomes:**

Small genome size (0.86-2.13 Mb) with a low GC percent and a proportion of coding genome ranging between 64.4% and 86%

3 genomic regions conserved among all the *Rickettsiales*, two operons encoding for ribosomal proteins and one region encoding for components of the type IV secretion system

Only *Anaplasma* spp. and *Ehrlichia* spp. same synteny across their chromosome.

OMP superfamily unique to the *Anaplasmataceae* family:

Over 100 members of OMP in *A. phagocytophilum*, 56 in *A. marginale*, 22 in *E. chaffeensis*, 16 in *E. ruminantium* and 25 in *E. canis*.

Organization of OMP cluster conserved among *Ehrlichia* spp.

Unique genes coding for membrane and regulatory proteins of *Rickettsiaceae*:

RickA and *sca2* genes, absent in the *Anaplasmataceae*, are present and functional in *Rickettsia* spp. Role in actin-based motility

Gene family *spoT* detected only in *Rickettsia* spp. genomes

Genome comparison between virulent and attenuated strains:

Characterization of genotypes associated with virulent phenotypes for *E. chaffeensis* and *R. prowazekii*

3) Functional genomics

3.1) General features

Transcriptomics and proteomics describe the complete, or nearly complete, collection of transcripts and proteins of an organism in different cells and tissues modulated in response to several conditions such as infectious processes. Careful analysis of the transcriptome and proteome is essential to understand the functional output of the genome [24].

More specifically, in addition to comparative genomic, they are powerful approaches to increase our knowledge on the *Rickettsiales* pathogenesis. Researchers have the opportunity to study 1/ the functional genomics of *Rickettsiales* in host or vector cells and 2/ the functional genomics of host or vector in response to *Rickettsiales* infection.

An overview of the different transcriptomic and proteomic studies currently available on the different *Rickettsiales* is represented in the figure 3. A complete analysis of host, vector and pathogen transcriptome was done solely for *A. phagocytophilum*. Globally, proteomic studies were performed only on the pathogen and its host: there is a lack of data concerning the proteome of the vector after infection. For *R. conorii* and *E. ruminantium*, transcriptomic and proteomic studies were performed only on the pathogen and there is no data on host or vector response to pathogen infection. Proteome analysis was done on same pathogens as transcriptomic studies with an additional species *R. felis* [25]. The use of such “Omics” approaches is particularly useful to perform integrative analyses, linking gene and protein expression for *A. marginale*, *A. phagocytophilum*, *E. chaffeensis*, *E. ruminantium*, *R. conorii* and *R. prowazekii*. The effect of *Wolbachia* on its host gene and protein expressions was studied (*Anopheles gambiae* model) in order to improve our comprehension of this symbiotic interaction.

Globally, most of the functional genomics studies were performed on the pathogen, only preliminary studies on the host side being available. Thus, there is a strong need to study both

vector and host transcriptome and proteome in response to infection in order to better understand pathogen-host and pathogen-vector interactions.

3.2) Functional genomics of *Rickettsiales*

Functional genomics studies of *Rickettsiales* are essential to identify main groups of genes regulated during infection of the host or vector. According to their COG, these genes are classified into the following categories: cell wall membrane biogenesis (membrane protein); translation, replication and post-translational modifications (PTM); amino acids biosynthesis, metabolism and transport; intra-cellular trafficking and secretion; energy production; related genes to virulence and pathogenesis; and the largest category comprise genes coding for hypothetical proteins.

3.2.1) Strategies and methods to perform transcriptomic and proteomic studies on *Rickettsiales*

Various strategies are used to perform *Rickettsiales* transcriptomic and proteomic studies. Authors generally compare the transcriptome and/or the proteome of i) bacteria infecting different host and/or vector cells, ii) bacterial strains with distinct virulence level and/or from geographically distinct regions, and iii) bacteria in *in vitro*, *in vivo* and *ex vivo* conditions. Until now, few transcriptomic studies were performed on the different stages of development of a *Rickettsiales* spp, mainly due to technical constraints associated to the intracellular properties of the bacteria. In one of these studies, the authors compared the *in vitro* gene expression profile between *E. ruminantium* elementary bodies and reticulate bodies (non-infectious forms) [26]. For *Wolbachia* spp., *E. chaffeensis*, *E. ruminantium*, *A. marginale*, *A. phagocytophilum*, *R. conorii*, *R. prowazekii* and *R. felis*, transcriptomic and proteomic studies were performed in their host cells (*in vitro*, *in vivo* or *ex vivo*). Additional studies on the pathogens were done in tick cells for *A. phagocytophilum* and *E. chaffeensis* (transcriptome)

and for *R. prowazekii*, *E. chaffeensis*, *A. phagocytophilum* and *A. marginale* (proteome). This allows a better understanding of interactions between bacteria and its host or vector, and the identification of genes that are essential for *Rickettsiales* adaptation in host and vector cells. One of the major constraints to study the transcriptome and proteome of obligate intracellular bacteria is the large excess of mRNA and proteins of host or vector origin. The easiest way to analyze the transcriptome and proteome of *Rickettsiales* is then to study the free living infectious form of the bacteria *i.e.* the elementary bodies or dense cored cells. The removal of eukaryotic RNA and prokaryotic rRNA contaminants is a prerequisite step for *Rickettsiales* transcriptomic studies. The depletion of eukaryotic RNA contaminant has been achieved mostly through the use of subtractive hybridization technologies. The classical approach developed by Ambion (MicrobEnrich Kit), consists in capturing oligonucleotides coupled with magnetic beads that hybridize to the 18S and 28S ribosomal RNA and the poly-adenylated 3' tail of eukaryotic mRNA. An additional method known as MicrobExpress (Ambion), using similar magnetic beads, allows the enrichment of the proportion of bacterial mRNA by capturing ribosomal and transfer RNA [27, 28]. More recently, Emboulé and co-workers developed an alternative method for the *Rickettsiales* *E. ruminantium*. This method allows both the removal of the eukaryotic RNA and prokaryotic rRNA contaminants and mRNA enrichment from *E. ruminantium* origin [29] and could be adapted for other *Rickettsiales*.

For proteomics, various methods have been used to enrich the amount of intracellular bacteria proteins. Biochemical fractionation based on differential density and size, on harsh detergent treatments dissolving differentially host cells on flow cytometry sorting, or on a combination of any of these methods has been used. Purity and yield of the enriched bacteria samples vary but generally, dozens to hundreds of bacteria proteins could be successfully detected [30].

3.3) Role of OMPs and post-translational modifications (PTM) during infection

3.3.1) In host cells:

The OMPs are detected in the proteome of 8 *Rickettsiales* studied [23, 25, 31-35]. They are the predominantly proteins detected from global proteome of these bacteria. For example, among the 113 *p44* paralogous genes in *A. phagocytophilum*, 110 of them were expressed in the human promyelocytic leukemia cell line, HL-60 [32]. The first proteome of *E. ruminantium* elementary bodies showed that MAP1 protein was found to be the most predominant protein and seems to be organized as a porin [33]. Other OMPs such as MAP 1-6, MAP 1-14 and MAP 1+1 were also identified. Pruneau and co-workers showed an up-regulation of *map1-6* gene expression for *E. ruminantium* reticulate bodies (intracellular forms) compared to elementary bodies (free living form) [26]. MAP protein family seems to be essential for *E. ruminantium* intracellular survival.

Genes encoding for PTMs were also found to be up-regulated during *Rickettsiales* infection. This is not fully surprising since PTMs have been detected in proteins involved in pathogenesis and virulence, allowing the pathogens to escape from host or vector immune response [36].

In *R. prowazekii*, genes encoding for OMPs were down-regulated in virulent strain compared to the avirulent strain, moreover, differences in methylation for OMPs were also observed between virulent and avirulent strains [23]. These results suggested the importance of OMPs and PTMs in pathogenesis. The down-regulation of genes encoding for OMP family proteins and the methylation of certain OMPs can allow escaping from host or vector immune response.

3.3.2) Comparison of OMP gene/protein expression in tick and host cells

In *A. phagocytophilum*, the locus *p44*, which encodes for several OMPs was transcribed in human cell lines and not in tick cell lines. Furthermore, the authors showed an important

proportion of up-regulated genes coding for others OMPs in *A. phagocytophilum* infecting human cell lines than in tick cell lines [37]. For *E. ruminantium*, MAP 1-1 was expressed in tick cells only and not in host cells [38]. A differential expression of OMP-family genes between host cells and vector cells was also observed for *E. chaffeensis* [39]. The OMPs of *Rickettsiales* are strongly regulated depending on host or vector cell types and could be important for the development within host and vector.

It would be interesting to better understand the role of these OMPs, which are at the interface between the pathogen and the host and vector and can represent potential vaccine candidates.

3.3.3) Differential PTM of proteins in tick and host cells

Rickettsiales with up-regulation of genes encoded for PTM can modify specific proteins, thus modulate host stress response and escape immune recognition. For *E. chaffeensis*, genes involved in PTM were differentially expressed between host and vector cells. Furthermore, Kuriakose and co-workers, showed a post-transcriptional regulation of certain *E. chaffeensis* genes involved in host-pathogen-vector interactions [39]. There is an increased interest to study PTM on bacteria associated with pathogenesis.

3.4) Identification of Type IV secretion systems (T4SS)

In a recent study, global proteomes in human promyelocytic leukemia cell line of *A. phagocytophilum* and *E. chaffeensis* were characterized [32]. Proteins of T4SS such as VirB4 and VirD4 were detected as well as the T4SS effectors AnkaA and Ats-1. T4SS components, namely VirB9 (basal body) and VirB11 (energy source), were detected by proteomic analysis of *E. ruminantium* [33]. For *A. phagocytophilum*, the up-regulation of these genes had already been demonstrated in 3 cell lines: in human cells (HL-60 and HMEC-1) and in tick cells (ISE6) [37]. Surprisingly, the *virB2* paralogs of the T4SS were differentially transcribed between the human and tick cells. This result can reflect a specific use of T4SS components

depending on the host. Moreover, the effector AnkA was strongly transcribed in HMEC-1, less transcribed in HL-60, and only marginally transcribed in ISE6 [37]. In another transcriptomic study, *ankA* was identified by RNA-sequencing in *Ixodes scapularis* tick salivary glands infected with *A. phagocytophilum* [40]. For *R. prowazekii*, the same components of T4SS were detected in global proteome of virulent and avirulent strains. It would appear therefore that there are no functional differences of the T4SS between virulent and avirulent strains [41]. Several components of T4SS were also detected in global proteome of endosymbiont *Wolbachia* [31]. This suggests a potential role of the T4SS during the endosymbiotic interaction.

3.5) Expression of proteins involved in the subversion of host immune response

In two transcriptomic studies, genes encoding for thioredoxin, *trx*, were found to be over-expressed for *E. ruminantium* [26] and for *R. conorii* [28]. The protein thioredoxin was also found in global proteome of virulent and avirulent *R. prowazekii* [41]. Oxidative stress is part innate immune response and results of a production of reactive oxygen species (ROS) from host cells to degrade the bacterial membrane. The induction of anti-oxidant enzymes such as thioredoxin or superoxide dismutase diminishes the ROS activity. Others genes involved in the defense against oxidative stress were also up-regulated for *E. chaffeensis* (ECH_0493 encodes for superoxide dismutase) [39] and for *A. phagocytophilum* (APH_0795 encodes for antioxidant AhpC/Tsa family) [37]. Superoxide dismutase was also detected in global proteome of *R. conorii* [35], *R. prowazekii* [41] and *E. chaffeensis* [42]. In *E. ruminantium* proteome, TsaA and ElbB proteins both involved in cell redox homeostasis were detected. Interestingly, amongst the several bacterial species compared, the authors found that ElbB was exclusively detected in *E. ruminantium* [33]. In host and vector cells, *Rickettsiales* also fight against osmotic stress by up-regulation of proline-betaine transporter (*proP*). Proline and betaine are two osmoprotectants which permit to overcome the inhibitory effects of

hyperosmolarity. The gene encoding for proP was up-regulated in *R. conorii* [28] and the protein was detected for the others *Rickettsiales* such as *R. prowazekii* [41]. These results reflect the successful adaptation of *Rickettsiales* to the host or vector cells fighting against the first defense mechanisms of cells.

3.6) Importance of hypothetical proteins during infection

Rickettsiales comparative genomic revealed a high percentage of genes encoding for hypothetical proteins (Table 2). Regardless to the adopted strategy in transcriptome or proteome studies, genes encoding for hypothetical proteins were one of the most represented COG. In a recent study, the authors compared the *E. chaffeensis* transcriptome in mammalian versus arthropod hosts [39]. They showed that genes encoding for hypothetical proteins were the most up-regulated genes in human and tick cells. Furthermore, among these genes, they identified 11 highly expressed in human cells that were not expressed in tick cells and 18 expressed in tick cells, not in human cells. These genes do not have any orthologs in other *Ehrlichia spp.* and seem to be required for adaptation and survival in host and vector cells [39]. Further studies are needed to characterize the function of these genes encoding for hypothetical proteins and their role in adaptation and survival in host or vector. More recently, two studies allowed characterizing the function of the hypothetical protein APH_1235 in *A. phagocytophilum*. The gene encoding for APH_1235, was found to be up-regulated in dense cored form [43] and the blocking of protein APH_1235 with antibodies reduced infection levels in mammalian cells [40]. It seems that APH_1235 is required for host cell infection. This protein has homologs in other *Anaplasma spp.* and *Ehrlichia spp.* but not in other bacteria. In *E. ruminantium*, CDS_00640, the ortholog of APH_1235 was also up-regulated in the infectious elementary bodies similar to the dense cored form in *A. phagocytophilum* [26]. Even if gene homology analysis is the first step to characterize unknown genes, further

experiments will be useful to validate the role of this *Ehrlichia ruminantium* gene in host cell infection.

3.7) Identification of pathogen metabolic activities in host and vector

Functional genomics of the 8 *Rickettsiales* focused in this review, revealed the presence of genes/proteins involved in metabolic pathways in host and vector cells. Indeed, many over-expressed genes or proteins were involved in energy production and conversion; nucleotide, amino acid, inorganic ion, carbohydrate and coenzyme transport and metabolism. Concerning amino acids biosynthesis, *Rickettsiales* members are able to synthesize few amino acids and must rely on transporting them from the host or vector cells. Within the *Rickettsiales*, the *Anaplasmataceae* members have very similar metabolic pathways which are distinct from those of the *Rickettsiaceae* members. *Wolbachia spp.* also differs from the other members of this family, in its inability to synthesize some cofactors [10]. For example, regarding the biotin biosynthesis (a key cofactors in *Rickettsiales*), it was observed that *E. chaffeensis* and *A. phagocytophilum* are able to produce biotin, but these genes were not expressed by *Wolbachia spp.* and *R. prowazekii* [10].

For *E. ruminantium*, transcriptomic study comparing different stages of development showed that genes involved in the carbohydrate, amino acid, inorganic ion, nucleotide, and coenzyme transports and metabolisms are differentially expressed at both reticulate and elementary body stages. These results suggest that elementary bodies of *E. ruminantium* could be metabolically active [26]. Moreover, at *E. ruminantium* elementary body stage, the majority of proteins detected are related to energy and general metabolism [33].

In *R. prowazekii*, proteins implicated in energy production and conversion; and coenzyme transport and metabolism, were more detected in avirulent strain than in virulent strain. This

result suggests that the differential abundance of these proteins may contribute to the differences in intracellular growth between the two strains [41].

Lin and co-workers analyzed global proteome of *A. phagocytophilum* and *E. chaffeensis*. They identified many proteins involved in nucleotide, vitamin, and cofactor biosynthetic pathway [32]. The comparison between *E. chaffeensis* transcriptome in human cells and in tick cells revealed a larger number of genes with high expression levels in the tick cells. These genes encoded for energy production and conversion, nutrient transport, metabolism, cellular process and translation. The up-regulation of these genes suggests that *E. chaffeensis* has higher metabolic activity in the vector cells [39].

These results suggest that *Anaplasmataceae* are not completely dependent on host or vector cells and may even supply vector cells with essential vitamins and nucleotides underlining the advantage of ticks to be infected. Moreover, metabolic activities of *Rickettsiales* seem to be related to both host or vector cells and different stages of development.

3.8) Functional genomics of host cells in response to infection

Regarding functional genomics of host cells in response to infection with *Rickettsiales* members, up-regulated genes are generally involved in defense mechanisms and immune responses such as genes encoding for interferon, cytokines, chemokines and their receptors.

Many of these genes were over-expressed in host cells infected with *A. phagocytophilum* [44] and *E. chaffeensis* [45]. These genes encode for proteins essential in the first defenses of host cells against bacteria. For *A. phagocytophilum*, the up-regulation of these genes was observed since the early times post-infection, most of them being down-regulated at later stages post-infection. It seems that *A. phagocytophilum* modulates the expression of these genes to survive in host cells [44].

Apoptotic pathways are affected resulting in an antiapoptotic state of host cells. This subversion of apoptotic pathways is due to the up-regulation of anti-apoptotic genes and down-regulation of pro-apoptotic genes. This was observed for *A. phagocytophilum* [44] and *E. chaffeensis* [45] but also for *R. prowazekii* [23].

In the study by Miura and Rikihisa, liver cell transcriptome was studied in response to infection with the three strains of *E. chaffeensis*: Wakulla, Liberty and Arkansas, which induce different histopathologic lesions in liver tissue. The expression profile of cytokines, chemokines and their receptors were found to be different among the three strains and could be therefore related to the distinct histopathologic lesions [45].

3.9) Functional genomics of vector cells in response to infection

Concerning vector response to *Rickettsiales* infection, 3 transcriptomics studies only were performed. In the first study, the authors compared *Ixodes scapularis* transcriptome in response to infection with *A. phagocytophilum* and *A. marginale* [46]. Gene expression profiles were different between the 2 *Anaplasma* species. This difference may reflect the difference of *Anaplasma spp.* developmental cycle in tick cells or that *I. scapularis* is not natural vector of *A. marginale* [46]. The 2 other studies have been done on the vector *Rhipicephalus microplus* infected with *A. marginale* [47, 48]. The authors showed that few genes were regulated in *R. microplus* salivary glands after infection with *A. marginale* [48]. There was thus limited impact of *A. marginale* infection on the tick gene expression compare to uninfected ticks suggesting minor effects on tick activity or survival. The differentially expressed genes encoding for putative proteins are involved in binding, catalytic/enzymatic activity, transport, DNA/RNA metabolism and structural molecules. Among them, 3 genes encoding for a putative von Willebrand factor, a flagelliform silk protein, and a subolesin genes, seem to be important for infection and multiplication of *A. marginale* in *R. microplus*.

559 In the third study, the authors showed a differential gene expression between the *R. microplus*
560 midgut and salivary gland in response to feeding and regulation in the salivary gland along
561 time [47]. This result illustrates the importance to study the pathogen-vector interaction
562 during the feeding process.

563 **Gap analysis on “omics” studies on *Rickettsiales*:**

Lack of study on the vector/host transcriptome and proteome after *Rickettsiales* infection

Few transcriptomic studies were done on the different stages of development of *Rickettsiales* spp.

Genes/proteins potentially involved in the pathogenesis:

Important role of OMPs and PTM in the development of *Rickettsiales* within host and vector and in escaping immune response

Ability of the pathogen to counteract the host innate responses: fight against oxidative and osmotic stress

Subversion of apoptotic pathways with over-expression of anti-apoptotic genes and down-regulation of pro-apoptotic genes

Genes/proteins involved in the pathogen development in ticks:

Limited impact of *A. marginale* infection on the tick gene expression compare to uninfected ticks suggesting minor effects on tick activity or survival

3 genes encoding for a putative von Willebrand factor, a flagelliform silk protein, and a subolesin genes, seem to be important for infection and multiplication of *A. marginale* in *R. microplus*

564

565 **4) Conclusion and future perspectives**

566 Comparative genomic allows identification of several important genes associated with strict
567 intracellular lifestyle, with host specificity and related to virulence phenotypes.

568 Complementary functional genomics permits to identify modulation of gene and protein
569 expression depending on environmental conditions. A first screening of gene/protein
570 expression by “Omics” approaches on the pathogen/host/vector allows, in a second time, the
571 focusing on specific mechanisms, gene function and pathway of interest to go further in the
572 comprehension of the pathogen biology and pathogenesis using classical approaches.

573 Resistance to host innate defense mechanisms like osmotic stress and oxidative burst are
574 identified in several studies and these mechanisms seem to play a key role for the
575 *Rickettsiales* intracellular development. Moreover, OMPs and PTM seem to have a key role in
576 the pathogenesis and their role should be further studied.

577 The major transcripts and proteins identified by these studies were associated with unknown
578 functions. They probably have an important role in pathogenesis and should be investigated.

579 Right now, the production of knock-out mutants for genes of interest is technically feasible in
580 *Rickettsiales*. The recent development of transient and stable *in vitro* transfection systems for
581 *A. marginale* [49, 50] and the optimization of random and targeted mutagenesis for *E.*

582 *chaffeensis* [51] pave the way for future exciting developments as they will allow the

583 functional analysis of these bacteria. Development of reliable high throughput RNA

584 sequencing methods which replace the use of microarrays is an amazing progress for the

585 understanding of the bacteria biology: all RNAs can be identified including small non-coding

586 RNAs. Up to now non-coding RNAs have not been studied in *Rickettsiales*. They could be

587 involved in regulation of key processes and could have a major role in the pathogenesis of

588 *Rickettsiales*.

In the past decade, significant efforts in improving analytical technologies relating to measuring mRNA, proteins and metabolites have been made. Nowadays, new breakthroughs in host-pathogen-vector research rely on the development of novel “Omics” approaches that incorporate high throughput sequencing or separation technologies, and other less known “omics” such as metabolomics, immunomics, and vaccinomics [52]. In the future, integrated “omics” investigation of various cellular molecules and their interaction in cells (i.e. interactomes) could lead to a quantified description of cellular metabolism for further hypothesis-driven investigation [53]. Those efforts will probably lead to fundamental new insights into bacterial metabolism during intracellular development.

5) Authors’ contributions

LP: Gathering information and papers, writing the outlines, writing the paper; NV: Proposing the subject, reviewing and editing the paper; AM, DFM: Preparation and interpretation of the pictures, gathering information and reviewing the paper. TL and IM: reviewing the paper. All authors agreed on outlines and the final version of the paper.

6) Acknowledgement

The authors acknowledge the financial support received from European project, FEDER 2007-2013, FED 1/1.4-30305, “Risque en santé animale et végétale” and the Fundação para a Ciência e Tecnologia (FCT, Lisbon, Portugal; contract number PTDC/CVT/114118/2009). LP and AM acknowledge financial support for their PhD from the European project, FED 1/1.4-30305”. IM acknowledges financial support from the grant SFRH/ BPD/ 45978/ 2008 from FCT. We thank Dr Jennifer Pradel and Claire Fritsch for their helpful corrections.

7) References

1. Merhej V, Raoult D: ***Rickettsial* evolution in the light of comparative genomics.** *Biol Rev Camb Philos Soc* 2011, **86**:379-405.
2. Dumler JS, Barbet AF, Bekker CP, Dasch GA, Palmer GH, Ray SC, Rikihisa Y, Rurangirwa FR: **Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*.** *Int J Syst Evol Microbiol* 2001, **51**:2145-2165.
3. Long SW, Zhang X, Zhang J, Ruble RP, Teel P, Yu XJ: **Evaluation of transovarial transmission and transmissibility of *Ehrlichia chaffeensis* (*Rickettsiales*: *Anaplasmataceae*) in *Amblyomma americanum* (Acari: Ixodidae).** *J Med Entomol* 2003, **40**:1000-1004.
4. Stich RW, Kocan KM, Palmer GH, Ewing SA, Hair JA, Barron SJ: **Transstadial and attempted transovarial transmission of *Anaplasma marginale* by *Dermacentor variabilis*.** *Am J Vet Res* 1989, **50**:1377-1380.
5. Weinert LA, Werren JH, Aebi A, Stone GN, Jiggins FM: **Evolution and diversity of *Rickettsia* bacteria.** *BMC Biol* 2009, **7**:6.
6. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, et al: **IMG: the Integrated Microbial Genomes database and comparative analysis system.** *Nucleic acids research* 2012, **40**:D115-122.
7. Georgiades K, Merhej V, El Karkouri K, Raoult D, Pontarotti P: **Gene gain and loss events in *Rickettsia* and *Orientia* species.** *Biol Direct* 2011, **6**:6.
8. Fournier PE, El Karkouri K, Leroy Q, Robert C, Giumelli B, Renesto P, Socolovschi C, Parola P, Audic S, Raoult D: **Analysis of the *Rickettsia africae* genome reveals that**

- virulence acquisition in *Rickettsia* species may be explained by genome reduction.** *BMC Genomics* 2009, **10**:166.
9. Frutos R, Viari A, Vachieri N, Boyer F, Martinez D: ***Ehrlichia ruminantium*: genomic and evolutionary features.** *Trends Parasitol* 2007, **23**:414-419.
 10. Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen JA, Seshadri R, Ren Q, Wu M, Utterback TR, et al: **Comparative genomics of emerging human ehrlichiosis agents.** *PLoS Genet* 2006, **2**:e21.
 11. Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y: **Functional diversity of ankyrin repeats in microbial proteins.** *Trends Microbiol* 2010, **18**:132-139.
 12. Wakeel A, Zhu B, Yu XJ, McBride JW: **New insights into molecular *Ehrlichia chaffeensis*-host interactions.** *Microbes Infect* 2010, **12**:337-345.
 13. Renvoise A, Merhej V, Georgiades K, Raoult D: **Intracellular *Rickettsiales*: Insights into manipulators of eukaryotic cells.** *Trends Mol Med* 2011, **17**:573-583.
 14. Nunez PA, Soria M, Farber MD: **The twin-arginine translocation pathway in alpha-proteobacteria is functionally preserved irrespective of genomic and regulatory divergence.** *PLoS One* 2012, **7**:e33605.
 15. Gillespie JJ, Brayton KA, Williams KP, Diaz MA, Brown WC, Azad AF, Sobral BW: **Phylogenomics reveals a diverse *Rickettsiales* type IV secretion system.** *Infect Immun* 2010, **78**:1809-1823.
 16. Rikihisa Y, Lin M: ***Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins.** *Curr Opin Microbiol* 2010, **13**:59-66.
 17. Lin J, Huang S, Zhang Q: **Outer membrane proteins: key players for bacterial adaptation in host niches.** *Microbes Infect* 2002, **4**:325-331.
 18. Achouak W, Heulin T, Pages JM: **Multiple facets of bacterial porins.** *FEMS Microbiol Lett* 2001, **199**:1-7.

19. Chan YG, Riley SP, Martinez JJ: **Adherence to and invasion of host cells by spotted Fever group *Rickettsia* species.** *Front Microbiol* 2010, **1**:139.
20. Gouin E, Egile C, Dehoux P, Villiers V, Adams J, Gertler F, Li R, Cossart P: **The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex.** *Nature* 2004, **427**:457-461.
21. Haglund CM, Choe JE, Skau CT, Kovar DR, Welch MD: ***Rickettsia* Sca2 is a bacterial formin-like mediator of actin-based motility.** *Nat Cell Biol* 2010, **12**:1057-1063.
22. Miura K, Rikihisa Y: **Virulence potential of *Ehrlichia chaffeensis* strains of distinct genome sequences.** *Infect Immun* 2007, **75**:3604-3613.
23. Bechah Y, El Karkouri K, Mediannikov O, Leroy Q, Pelletier N, Robert C, Medigue C, Mege JL, Raoult D: **Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities.** *Genome Res* 2010, **20**:655-663.
24. Filiatrault MJ: **Progress in prokaryotic transcriptomics.** *Curr Opin Microbiol* 2011, **14**:579-586.
25. Ogawa M, Renesto P, Azza S, Moinier D, Fourquet P, Gorvel JP, Raoult D: **Proteome analysis of *Rickettsia felis* highlights the expression profile of intracellular bacteria.** *Proteomics* 2007, **7**:1232-1248.
26. Pruneau L, Emboule L, Gely P, Marcelino I, Mari B, Pinarello V, Sheikboudou C, Martinez D, Daigle F, Lefrancois T, et al: **Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development.** *FEMS Immunol Med Microbiol* 2012, **64**:66-73.
27. La MV, Francois P, Rovey C, Robineau S, Barbry P, Schrenzel J, Raoult D, Renesto P: **Development of a method for recovering *rickettsial* RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria.** *J Microbiol Methods* 2007, **71**:292-297.

28. Renesto P, Rovey C, Schrenzel J, Leroy Q, Huyghe A, Li W, Lepidi H, Francois P, Raoult D: ***Rickettsia conorii* transcriptional response within inoculation eschar.** *PLoS One* 2008, **3**:e3681.
29. Emboule L, Daigle F, Meyer DF, Mari B, Pinarello V, Sheikboudou C, Magnone V, Frutos R, Viari A, Barbry P, et al: **Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*.** *BMC Mol Biol* 2009, **10**:111.
30. Marcelino I, de Almeida AM, Ventosa M, Pruneau L, Meyer DF, Martinez D, Lefrancois T, Vachery N, Coelho AV: **Tick-borne diseases in cattle: Applications of proteomics to develop new generation vaccines.** *J Proteomics* 2012.
31. Bennuru S, Meng Z, Ribeiro JM, Semnani RT, Ghedin E, Chan K, Lucas DA, Veenstra TD, Nutman TB: **Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*.** *Proc Natl Acad Sci U S A* 2011, **108**:9649-9654.
32. Lin M, Kikuchi T, Brewer HM, Norbeck AD, Rikihisa Y: **Global proteomic analysis of two tick-borne emerging zoonotic agents: *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*.** *Front Microbiol* 2011, **2**:24.
33. Marcelino I, de Almeida AM, Brito C, Meyer DF, Barreto M, Sheikboudou C, Franco CF, Martinez D, Lefrancois T, Vachery N, et al: **Proteomic analyses of *Ehrlichia ruminantium* highlight differential expression of MAP1-family proteins.** *Vet Microbiol* 2012, **156**:305-314.
34. Ramabu SS, Ueti MW, Brayton KA, Baszler TV, Palmer GH: **Identification of *Anaplasma marginale* proteins specifically upregulated during colonization of the tick vector.** *Infect Immun* 2010, **78**:3047-3052.

35. Renesto P, Azza S, Dolla A, Fourquet P, Vestris G, Gorvel JP, Raoult D: **Proteome analysis of *Rickettsia conorii* by two-dimensional gel electrophoresis coupled with mass spectrometry.** *FEMS Microbiol Lett* 2005, **245**:231-238.
36. Ribet D, Cossart P: **Post-translational modifications in host cells during bacterial infection.** *FEBS Lett* 2010, **584**:2748-2758.
37. Nelson CM, Herron MJ, Felsheim RF, Schloeder BR, Grindle SM, Chavez AO, Kurtti TJ, Munderloh UG: **Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis.** *BMC Genomics* 2008, **9**:364.
38. Postigo M, Taoufik A, Bell-Sakyi L, Bekker CP, de Vries E, Morrison WI, Jongejan F: **Host cell-specific protein expression in vitro in *Ehrlichia ruminantium*.** *Vet Microbiol* 2008, **128**:136-147.
39. Kuriakose JA, Miyashiro S, Luo T, Zhu B, McBride JW: ***Ehrlichia chaffeensis* transcriptome in mammalian and arthropod hosts reveals differential gene expression and post transcriptional regulation.** *PLoS One* 2011, **6**:e24136.
40. Mastronunzio JE, Kurscheid S, Fikrig E: **Postgenomic analyses reveal development of infectious *Anaplasma phagocytophilum* during transmission from ticks to mice.** *J Bacteriol* 2012, **194**:2238-2247.
41. Chao CC, Chelius D, Zhang T, Mutumanje E, Ching WM: **Insight into the virulence of *Rickettsia prowazekii* by proteomic analysis and comparison with an avirulent strain.** *Biochim Biophys Acta* 2007, **1774**:373-381.
42. Seo GM, Cheng C, Tomich J, Ganta RR: **Total, membrane, and immunogenic proteomes of macrophage- and tick cell-derived *Ehrlichia chaffeensis* evaluated by liquid chromatography-tandem mass spectrometry and MALDI-TOF methods.** *Infect Immun* 2008, **76**:4823-4832.

43. Troese MJ, Kahlon A, Ragland SA, Ottens AK, Ojogun N, Nelson KT, Walker NJ, Borjesson DL, Carlyon JA: **Proteomic analysis of *Anaplasma phagocytophilum* during infection of human myeloid cells identifies a protein that is pronouncedly upregulated on the infectious dense-cored cell.** *Infect Immun* 2011, **79**:4696-4707.
44. Lee HC, Kioi M, Han J, Puri RK, Goodman JL: ***Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells.** *Genomics* 2008, **92**:144-151.
45. Miura K, Rikihisa Y: **Liver transcriptome profiles associated with strain-specific *Ehrlichia chaffeensis*-induced hepatitis in SCID mice.** *Infect Immun* 2009, **77**:245-254.
46. Zivkovic Z, Blouin EF, Manzano-Roman R, Almazan C, Naranjo V, Massung RF, Jongejan F, Kocan KM, de la Fuente J: ***Anaplasma phagocytophilum* and *Anaplasma marginale* elicit different gene expression responses in cultured tick cells.** *Comp Funct Genomics* 2009:705034.
47. Mercado-Curiel RF, Palmer GH, Guerrero FD, Brayton KA: **Temporal characterisation of the organ-specific *Rhipicephalus microplus* transcriptional response to *Anaplasma marginale* infection.** *Int J Parasitol* 2011, **41**:851-860.
48. Zivkovic Z, Esteves E, Almazan C, Daffre S, Nijhof AM, Kocan KM, Jongejan F, de la Fuente J: **Differential expression of genes in salivary glands of male *Rhipicephalus (Boophilus) microplus* in response to infection with *Anaplasma marginale*.** *BMC Genomics* 2010, **11**:186.
49. Felsheim RF, Chavez AS, Palmer GH, Crosby L, Barbet AF, Kurtti TJ, Munderloh UG: **Transformation of *Anaplasma marginale*.** *Vet Parasitol* 2010, **167**:167-174.
50. Noh SM, Ueti MW, Palmer GH, Munderloh UG, Felsheim RF, Brayton KA: **Stability and tick transmission phenotype of gfp-transformed *Anaplasma marginale* through a complete in vivo infection cycle.** *Appl Environ Microbiol* 2011, **77**:330-334.

51. Cheng C, Nair AD, Indukuri VV, Gong S, Felsheim RF, Jaworski D, Munderloh UG, Ganta RR: **Targeted and Random Mutagenesis of *Ehrlichia chaffeensis* for the Identification of Genes Required for In vivo Infection.** *PLoS Pathog* 2013, **9**:e1003171.
52. Bagnoli F, Baudner B, Mishra RP, Bartolini E, Fiaschi L, Mariotti P, Nardi-Dei V, Boucher P, Rappuoli R: **Designing the next generation of vaccines for global public health.** *Omics* 2011, **15**:545-566.
53. Schilling CH, Schuster S, Palsson BO, Heinrich R: **Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era.** *Biotechnol Prog* 1999, **15**:296-303.
54. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.

Figure 1: Classification of *Rickettsiales* using 16S rDNA. The 31 16S rDNA sequences were aligned using MUSCLE [54]. Phylogenetic tree was constructed with ML method and rooted with an outgroup of others α -*proteobacteria*. One hundred bootstrap replicates were completed. Bootstrap values are indicated in percentage. Different colors correspond to different species for *Anaplasmataceae* family and to different groups for *Rickettsia* spp.

Figure 2. Intracellular lifestyles of pathogenic *Rickettsiaceae* and *Anaplasmataceae*. Intracellular pathogenic *Rickettsiales* invade host cell through endocytosis. After internalization, the bacteria can either persist in an intracellular vacuole (morulae) that is derived from host cell or ER membranes, or escape to the cytosol. (A) Bacteria of the *Rickettsiaceae* family escape from the phagosome and form an actin tail. The actin-based motility allows swift, unidirectional movement across the cytoplasm into adjacent cells, promoting cell-to-cell spread. (B) For pathogenic *Anaplasmataceae*, the dense-cored cell develops into replicating reticulate cells in the phagosome, which does not fuse with lysosome and forms morulae. Then, reticulate cells mature into dense-cored cells that are liberated by exocytosis or host cell lysis. Arrows, transition between stages; bar-headed line, inhibition of phagolysosomal fusion; ER, endoplasmic reticulum.

Figure 3: Overview of pathogen, host and vector transcriptomic and proteomic studies

Transcriptomic or proteomic studies done on host (pink circle) and on vector (dark blue circle) after pathogen infection (the name of the pathogen is indicated for each study). Transcriptomic or proteomic studies done on the pathogen (light blue). Intersections of circles correspond to studies done on the pathogen and on host or/and vector. In example, transcriptomic studies of *Anaplasma marginale* and of its vector have been done whereas only proteomic studies of *Anaplasma marginale* has been performed.

Table 1: The 8 *Rickettsiales* spp. selected, their hosts and vectors, diseases and geographical distribution

Species	Host	Vector	Disease	Geographical distribution
<i>A. phagocytophilum</i>	Humans, deer, dogs, cats, horses, ruminants, rodents	Ticks <i>Ixodes</i> spp.	Human granulocytic anaplasmosis, tick-born fever of ruminants, anaplasmosis	USA, Europe and Asia
<i>A. marginale</i>	Cattle, wild ruminants	Ticks <i>Rhipicephalus microplus</i> , <i>Dermacentor</i>	Bovine anaplasmosis	Worldwide in tropical and subtropical regions
<i>E. chaffeensis</i>	Human, deer, dogs	Ticks <i>Amblyomma americanum</i>	Human monocytic ehrlichiosis, canine ehrlichiosis	USA, South America and Asia
<i>E. ruminantium</i>	Cattle, sheep, goats, wild ruminants	Ticks <i>Amblyomma variegatum</i> and <i>hebraeum</i>	Heartwater	Sub Saharan Africa, Comoros, Mayotte, Reunion island, Madagascar, Caribbean
<i>Wolbachia</i> spp	Nematod <i>Brugia malayi</i> , Arthropod <i>Anopheles gambiae</i>	NA	NA	Worldwide distribution
<i>R. prowazekii</i>	Human, flying squirrels	Human lice <i>Pediculus humanus humanus</i>	Epidemic typhus	Worldwide distribution
<i>R. conorii</i>	Human, rodents	Ticks <i>Rhipicephalus sanguineus</i>	Mediterranean spotted fever	Mediterranean area and Africa
<i>R. felis</i>	Human, cats, dogs, rodents	Cat flea <i>Ctenocephalides</i>	Flea-borne spotted fever	North and South America, Southern Europ, Thailand and Australia

Table 2: Main genome features of the 8 *Rickettsiales* spp.

Family	<i>Anaplasmataceae</i>				<i>Rickettsiaceae</i>			
Species	<i>A.phagocytophilum</i>	<i>A. marginale</i>	<i>E. chaffeensis</i>	<i>E.ruminantium</i>	<i>Wolbachia</i> *	<i>R.prowazekii</i>	<i>R.conorii</i>	<i>R.felis</i>
Genome Size (bp)	1,471,282	1,197,687	1,176,248	1,499,920	1,267,782	1,111,523	1,268,755	1,587,240
Coding DNA (%)	72.2	86.0	79.7	64.4	85.7	76.12	80.93	83.89
GC (%)	41.6	49.8	30.1	27.5	35.2	29	32.4	32.5
ORFs	1,369	965	1,115	950	1,271	886	1,376	1,512
Hypothetical proteins	458 (33.5%)	151 (15.6%)	314 (28.2%)	401 (42.2%)	337 (26.5%)	298 (33.6%)	672 (48.8%)	476 (31.5%)
(% of ORFs)								
tRNAs	37	37	37	36	34	33	33	33
rRNAs	3	3	3	3	3	3	3	3

*:endosymbiont of arthropods and nematods

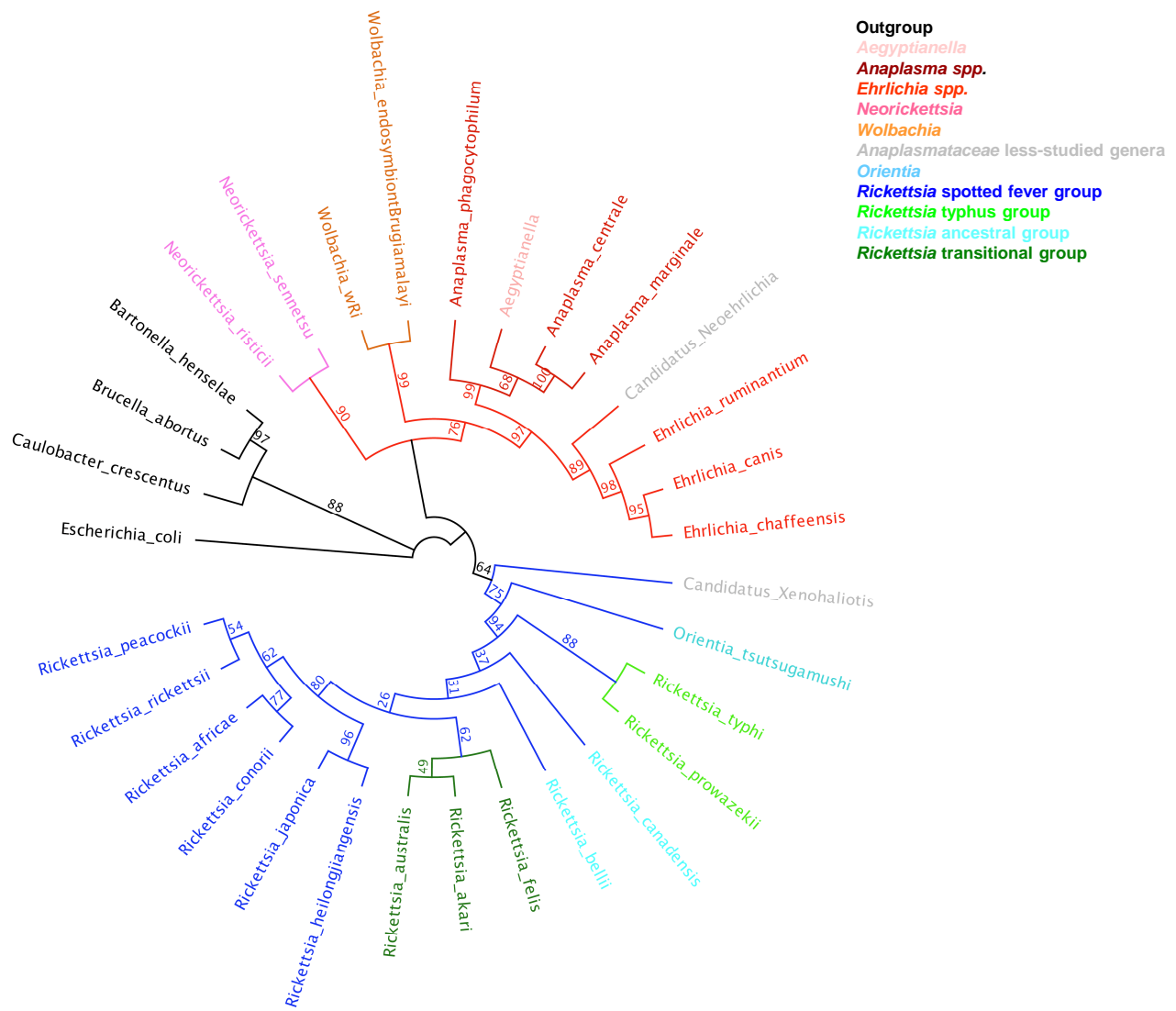


Figure 1

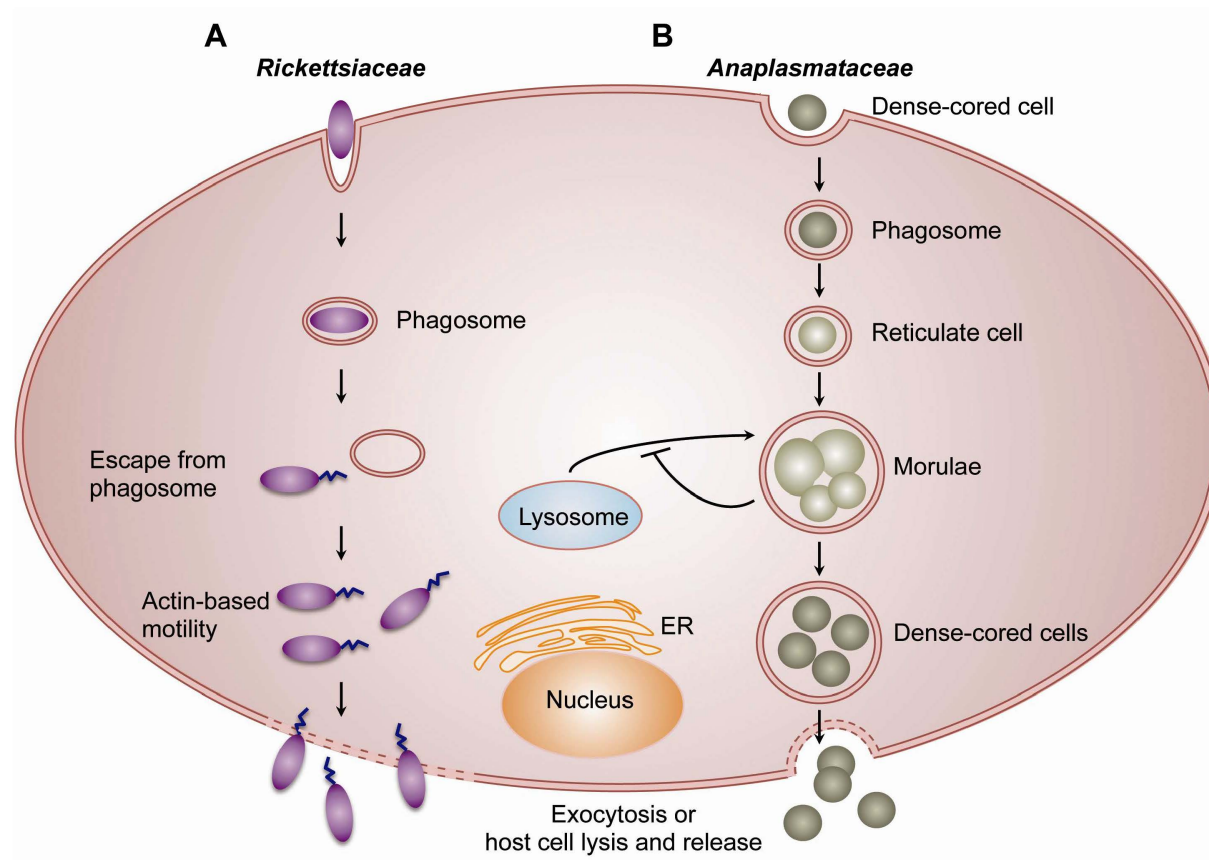
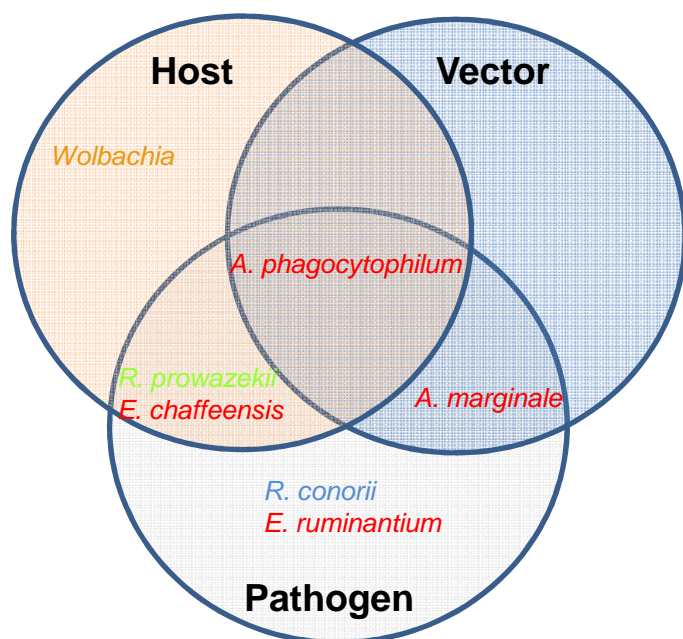


Figure 2

Transcriptomics



Proteomics

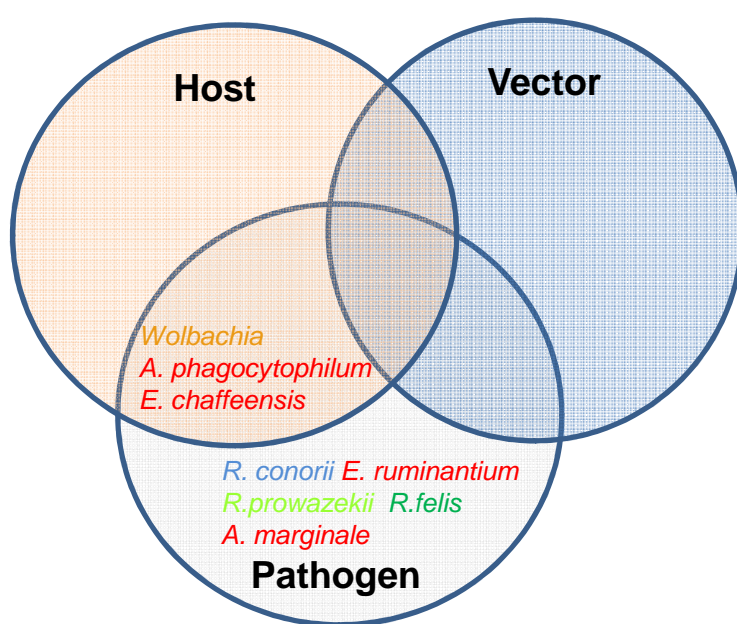


Figure 3

II. Chapitre 2 :

Stratégie d'étude et protocoles expérimentaux mis en œuvre pour l'analyse du transcriptome d'*Ehrlichia ruminantium*

SOMMAIRE

1.	Production des répliquats biologiques Gardel et Sénégal.....	88
1.1.	Isolement de Sénégal virulente à partir de chèvres infectées	88
1.2.	Synchronisation de l'infection et protocole de collecte des ARNs totaux.....	88
1.3.	Colorations et observations microscopiques aux différents temps post- infection.....	90
1.4.	Quantification bactérienne par qPCR.....	91
2.	Méthode SCOTS	91
2.1.	Protocole expérimental de la méthode SCOTS	92
2.2.	Vérification des captures SCOTS	93
3.	Analyse du transcriptome d' <i>E. ruminantium</i> par microarrays.....	94
3.1.	Principe des microarrays.....	94
3.2.	Design des puces à ADN spécifiques d' <i>E. ruminantium</i>	95
3.3.	Marquage, hybridation et analyse des puces à ADN.....	96
3.4.	Analyses statistiques des résultats.....	97
3.5.	Comparaison possibles des résultats.....	98
3.6.	Annotation des gènes différentiellement exprimés.....	99
3.7.	Validation de certains gènes par q-RTPCR	100

1. Production des réplicats biologiques Gardel et Sénégal

Pour l'analyse du transcriptome, l'ARNm d'*E. ruminantium* est produit à partir de cultures des souches Gardel et Sénégal virulente et atténuée sur les cellules endothéliales aortiques bovines (BAE). Les cultures sont effectuées en boîtes de culture Falcon 25 cm²: 2 boîtes sont utilisées à chaque point de collecte. Le milieu de culture utilisé est le Glasgow minimal essential medium additionné de L-glutamine (1%), de pénicilline-streptomycine (1%) et de *Tryptose phosphate broth* (TPB). Juste avant utilisation, il est enrichi en sérum de veau fœtal (10%) et les cellules sont incubées à 37 °C dans une atmosphère contenant 5% de CO₂.

1.1. Isolement de Sénégal virulente à partir de chèvres infectées

Un isolement de la souche Sénégal à partir d'un aliquot de sang infecté a été réalisé sur chèvres. Deux chèvres naïves ont donc été infectées expérimentalement par des cryotubes de sang contenant la souche Sénégal virulente. Un suivi journalier de la température et des signes cliniques a été réalisé jusqu'à l'apparition de l'hyperthermie qui, coïncide avec la circulation des CE d'*E. ruminantium* dans le sang. A ce moment, des prélèvements sanguins ont été effectués pendant les 2 jours de l'hyperthermie. Le sang prélevé dilué au 1/2 en milieu de culture a été incubé sur un tapis cellulaire de BAE. Les plages de lyse ont commencé à apparaître après une quinzaine de jours post infection.

1.2. Synchronisation de l'infection et protocole de collecte des ARNs totaux

La première étape a consisté à synchroniser les cultures des souches Gardel virulente et atténuée et Sénégal virulente et atténuée afin d'obtenir les ARN messagers correspondant à un même stade de développement. Un inoculum infectieux calibré a été utilisé pour infecter les TC25cm². Des collectes à plusieurs temps post infection devant correspondre aux

différents stades de développement ont été réalisées, cependant suivant les souches ces temps post infection étaient variables. Pour Gardel virulente qui présente une lyse complète en 5 jours, 5 points de collecte des échantillons ont été réalisés afin de récupérer tous les stades de développement : soit 24 hpi, 48 hpi, 72 hpi, 96 hpi et 120 hpi. Pour les souches Gardel et Sénégal atténuées qui présentent une lyse complète en 4 jours, 4 points de collecte (24 hpi, 48 hpi, 72 hpi et 96 hpi) ont été réalisés. Enfin, pour la souche Sénégal virulente qui présente une lyse complète en 6 jours, 6 points de collecte ont été réalisés (24 hpi, 48 hpi, 72 hpi, 96 hpi, 120 hpi, et 144 hpi). Pour l'ensemble des points de collecte, 1/10 du surnageant contenant les cellules infectées est récolté et conservé en culot sec pour l'extraction de l'ADN génomique permettant ainsi une estimation du nombre de copie d'*E. ruminantium* dans chaque échantillon par PCR quantitative en temps réel (qPCR) (figure 5).

Aux différents temps post-infection avant la lyse complète, les cellules endothéliales sont encore indemnes et les bactéries sont toujours à l'intérieur de la cellule hôte. Les cellules infectées sont trypsinées et rapidement lavées avec du milieu de culture d'infection pour arrêter l'action protéasique de la trypsine. Le surnageant contenant les cellules infectées est centrifugé à 1700 g pendant 5 min à 4°C et le culot cellulaire est rapidement lysé dans du TRIzol (Invitrogen, 2,5 ml pour une Tc 25 cm²) et conservé à -80°C pour extraire l'ARN total. Aux différents temps de lyse complète, le surnageant est collecté avec le tapis cellulaire préalablement trypsiné et ultracentrifugé pendant 15 min à 20 000g et à une température de 4°C. L'ensemble est lysé à l'aide du TRIzol et les échantillons sont conservés à -80°C avant l'extraction d'ARN (figure 5).

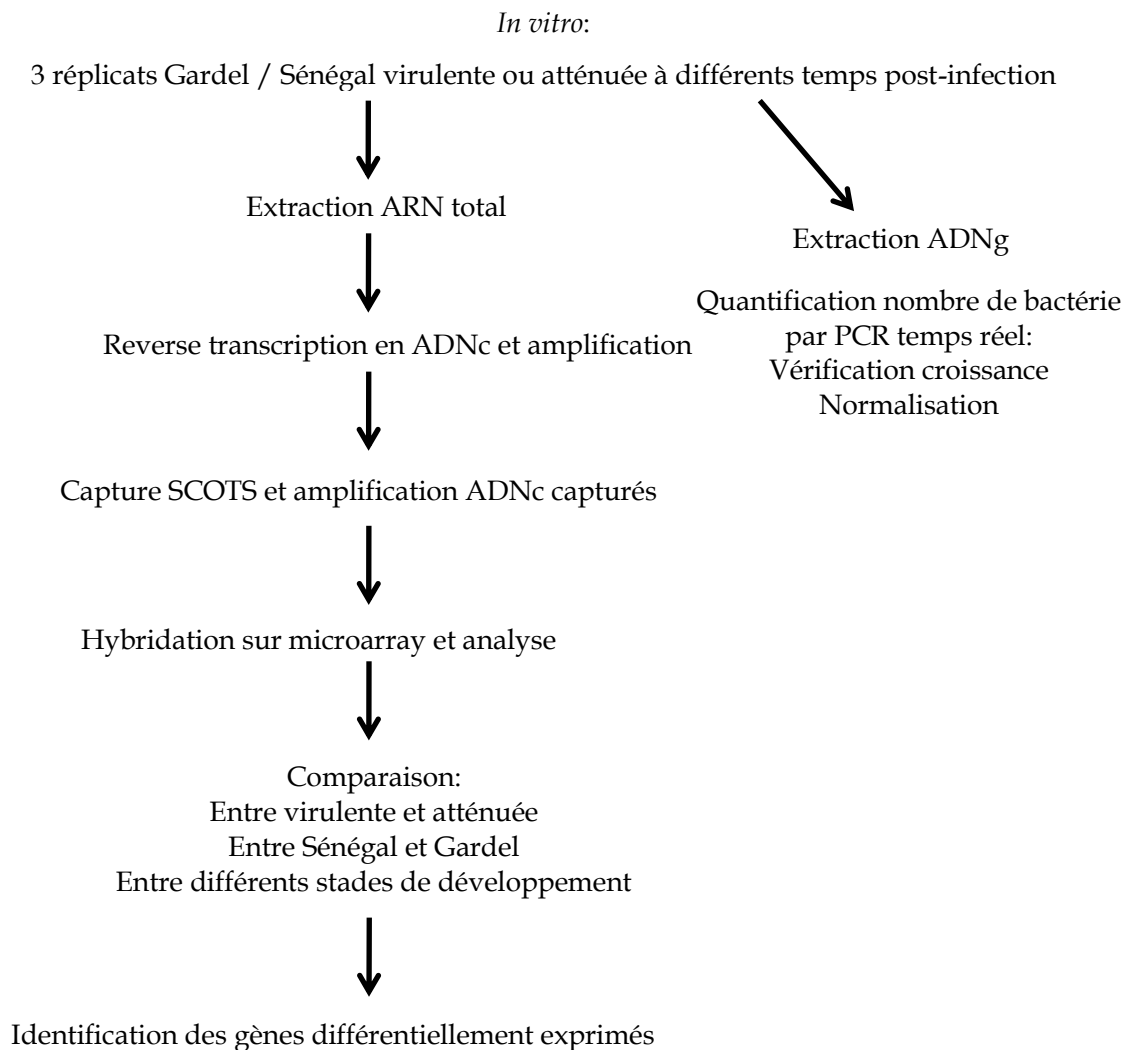


Figure 5 : Schéma générale de la stratégie expérimentale.

1.3. Colorations et observations microscopiques aux différents temps post-infection

Pour vérifier la synchronisation du développement d'*E. ruminantium*, une coloration RAL 555 est effectuée (fixation à l'Ethanol, coloration à l'Eosine et au Bleu de méthylène). Une observation journalière (avant et après coloration) par microscopie inversée des stades de développement de la bactérie est effectuée en parallèle. L'éosine colore en rose léger le

cytoplasme des cellules endothéliales tandis que les morulas et les CE d'*E. ruminantium* sont colorés en bleu violacé par le bleu de méthylène.

1.4. Quantification bactérienne par qPCR

A chaque point de collecte, on effectue une quantification de la bactérie à partir de l'ADNg extrait, par qPCR utilisant une sonde spécifique ciblant le gène *pCS20* spécifique d'*E. ruminantium*. La quantification est réalisée dans un volume réactionnel de 25 µL. Le mix est composé du Taqman universal PCR master mix (*Applied Biosystems*), de 900 nM d'amorces sens et anti-sens, de 250 nM de sonde et d'H₂O avec comme matrice 4 µL de l'ADNg extrait ([Martinez, et al., 2004](#)). De l'ADNg pur d'*E. ruminantium* de concentration connue est utilisé pour la réalisation de la gamme étalon obtenue à partir de dilutions successives. Un témoin négatif constitué d'eau est inclus pour chaque essai. Les conditions d'amplification dans le thermocycleur (ABI-Prism 7500, *Applied Biosystems*) sont : 2' à 50°C, 10' à 95°C, 40 cycles comprenant 15'' à 95°C, 1' à 60°C. La lecture de la fluorescence se fait à chaque cycle lors de la deuxième étape.

Les amorces utilisées ont les séquences suivantes:

- pCS20gardEx3F : 5' GGAGGATTAAAAGCAGCAAAAGACTT 3'

- pCS20gardEx3R : 5' GTTTTGCATCTTGTGGTGGTACTT 3'

La sonde *pCS20* spécifique d'*E. ruminantium* possède la séquence suivante :

- pCS20gardex3M2 : 5' 6-FAM-ATATGTGTTGCAGAATTT- BHQ1

Pour tous les échantillons, la qPCR est réalisée en duplicat.

2. Méthode SCOTS

L'utilisation des puces à ADN nécessite de l'ARNm de qualité et en quantité suffisante. Or l'isolement de l'ARNm d'*E. ruminantium*, présent en faible quantité parmi l'ARN total (eucaryote et procaryote) représente une phase critique. De plus, l'ARNm bactérien ne représente qu'une très faible

proportion de l'ARN total bactérien car l'ARNr est présent en grande quantité (> 80%). Pour pallier à ce problème nous avons eu recours à la méthode SCOTS initialement développée pour déterminer les gènes exprimés chez *Mycobacterium tuberculosis* lors de l'infection de macrophages humains en culture *in vitro*. Cette technique qui consiste à capturer les transcrits bactériens parmi l'ARN total a été adaptée avec succès à *E. ruminantium* ([Emboule, et al., 2009](#)).

2.1. Protocole expérimental de la méthode SCOTS

D'une façon générale, cette technique repose sur la capture spécifique des transcrits bactériens (Figure 6). L'ARN total du pathogène est d'abord isolé et converti en ADNc en utilisant des amorces spéciales. La partie 3' est un nonamère aléatoire (N9) tandis que la partie 5' est une séquence « étiquette » connue qui n'a pas d'homologie dans le génome des espèces étudiées. Ainsi, la conversion de l'ARN en ADNc se fait de façon aléatoire et ajoute la séquence étiquette à chacun des transcrits. Les transcrits peuvent ensuite être amplifiés par PCR en utilisant comme amorce la séquence étiquette. Une fois l'ADNc synthétisé, il s'agit de capturer les transcrits bactériens et de se débarrasser des transcrits eucaryotes et des transcrits ARN ribosomiaux bactériens (ARNr). Pour cela, l'ADNg de la bactérie est biotinylé et pré-hybridé avec un excès d'ADN ribosomal cloné bloquant ainsi les sites d'hybridation à l'ARNr. Vient ensuite l'étape de capture des transcrits, où l'ADNg ainsi préparé est hybridé avec les ADNc provenant du pathogène. Seuls les ADNc bactériens pourront s'hybrider à l'ADNg. Les ADNc codant pour l'ARNr ne seront pas capturés puisque les zones homologues sur l'ADNg sont déjà bloquées par l'ADNr. Les hybrides ADNc-ADNg sont récupérés à l'aide de billes magnétiques couvertes de streptavidine. Ensuite, les ADNc sont élués et amplifiés par PCR. Une autre ronde de capture est effectuée pour augmenter la quantité de transcrits spécifiques du pathogène et diminuer les contaminants (transcrits

eucaryotes et ARNr) (Faucher, 2007). Les ADNc ainsi obtenus, peuvent être hybridés aux puces à ADN.

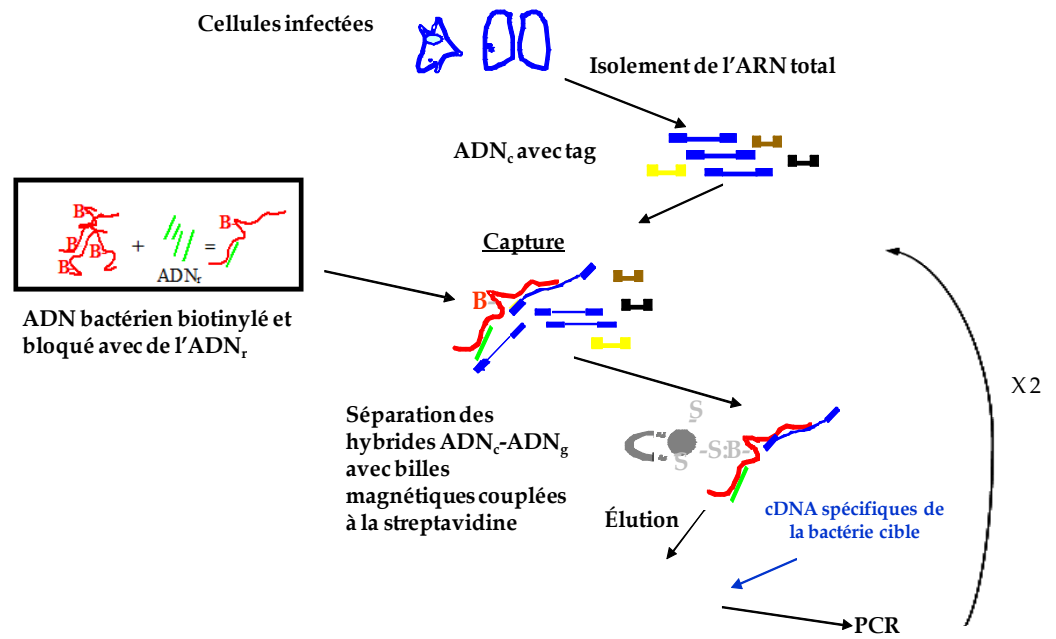


Figure 6: Méthode de capture sélective des transcrits ARNm (SCOTS) (Daigle, *et al.*, 2001)

2.2. Vérification des captures SCOTS

Afin de vérifier l'efficacité des captures SCOTS réalisées, une qPCR ciblant l'ARN 16S est effectuée. Cela permet de vérifier la diminution en transcrit 16S au cours des captures et donc l'efficacité de la méthode SCOTS. La quantification est réalisée dans un volume réactionnel de 25 μ L. Le mix est composé du Sybergreen fluorescent master mix reagent (*Applied Biosystem*), de 300 nM d'amorces sens et anti-sens, et d' H_2O avec comme matrice 2 μ L de l'échantillon SCOTS à tester. De l'ADN_g pur et de concentration connue est utilisé pour la réalisation de la gamme étalon obtenue à partir de dilutions successives. Un témoin négatif constitué d'eau est inclus pour chaque essai. Les conditions d'amplification dans le

thermocycleur (ABI-Prism 7500, *Applied Biosystems*) sont : 10' à 95°C, 40 cycles comprenant 30'' à 95°C et 1' à 60°C. La lecture de la fluorescence se fait à chaque cycle lors de la deuxième étape. A la fin des 40 cycles, une étape de dissociation est effectuée comme suit : 15'' à 95°C, 20'' à 60°C puis 15'' à 95°C.

Les amorces utilisées ont les séquences suivantes:

- 16-F1 : 5' AGCGCAACCCTCATCCTTAG 3'

- 16S-R1 : 5' AGCCCACCCTATAAGGGCC 3'

Pour tous les échantillons, la qPCR est réalisé en duplicat.

3. Analyse du transcriptome d'*E. ruminantium* par microarrays

3.1. Principe des microarrays

Les récentes avancées en biologie moléculaire et en biotechnologie, ont permis le développement de nouvelles techniques d'analyse d'expression génomique telles que les microarrays (figure 7). Aujourd'hui les microarrays sont utilisés comme outils de choix pour obtenir le profil d'expression global d'un pathogène dans des conditions expérimentales différentes.

L'analyse repose sur la mesure relative de l'expression des gènes entre deux conditions. Il s'agit de greffer sur une lame de verre des fragments « synthétiques » d'ADN (les sondes) représentatifs de chacun des gènes étudiés pour un organisme donné ([Ramsay, 1998](#), [Rockett, et al., 2001](#)). Les différentes sondes synthétisées, sont ensuite déposées sur le support (verre, silicium, plastique) sur lequel, les transcrits obtenus de la condition test et de la condition contrôle, et préalablement marqués par des fluorochromes différents, sont hybridés. Généralement, les fluorochromes utilisés correspondent aux Cyanine 3 et 5 (Cy3 et Cy5) (hybridation compétitive). On peut aussi hybrider séparément des transcrits obtenus de la condition test et comparer ainsi plusieurs conditions hybridées indépendamment

entre elles à l'aide d'un seul et même fluorochrome, en général le Cy3 qui est plus stable (Figure 7). Après hybridation des transcrits à chaque sonde, la mesure des intensités des signaux de fluorescence relatifs à chacun des fluorochromes permet finalement de calculer pour chaque gène le rapport des intensités, qui évalue son expression différentielle.

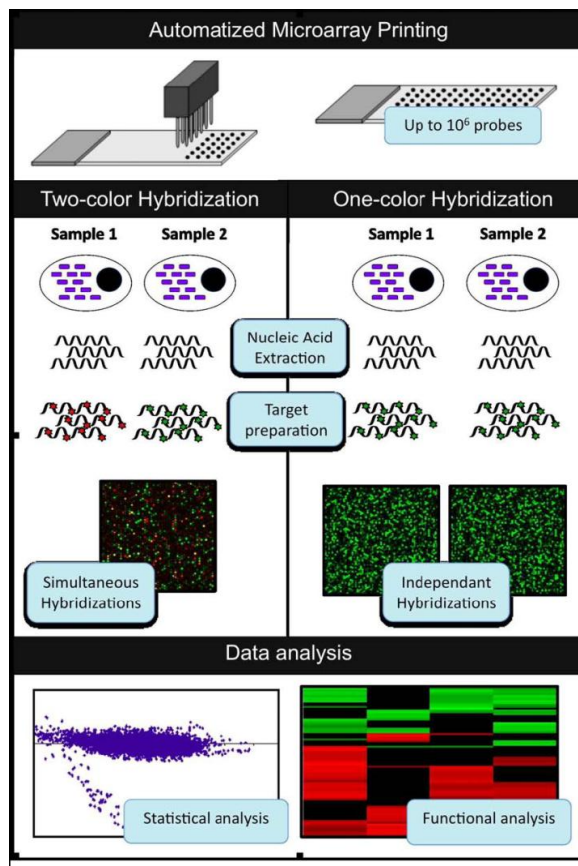


Figure 7: Schéma représentatif de la technologie des microarrays (www-lmgm.biotoul.fr)

3.2. Design des puces à ADN spécifiques d'*E. ruminantium*

Pour l'analyse du transcriptome de la souche Gardel virulente au cours du développement, les puces à ADN utilisées, ont été dessinées au cours de la précédente thèse. Le design de ces puces, les conditions de marquage et d'hybridation des ADNc, l'acquisition des données ainsi que la

normalisation des résultats sont décrits dans l'article présentant l'adaptation de la méthode SCOTS à *E. ruminantium* ([Emboule, et al., 2009](#)). Il s'agit de puces Agilent 8 x 15 k, c'est-à-dire des puces présentant 8 chambres d'hybridation et un maximum de 15 000 sondes spottées par chambre. Deux sondes ont été dessinées par CDS et spottées en 5 réplicats.

Les nouvelles puces dessinées au cours de ma thèse, ont été réalisées à partir des génomes des 6 souches d'*E. ruminantium* (Gardel et Sénégal virulente et atténuée, et Welgevonden Erwe et Erwo). Ce sont des puces Agilent 8 x 60 k, permettant ainsi de spotter jusqu'à 60 000 sondes par chambre d'hybridation. Nous avons dessiné deux types de sondes : des sondes dites « génériques » qui sont communes à l'ensemble des 6 souches, et des sondes dites « spécifiques » d'une ou plusieurs souches. Le design des sondes a été effectué à l'aide du logiciel Oligoarray par nos collaborateurs Kévin Lebrigand (ingénieur bioinformaticien) et Bernard Mari (chercheur) de l'institut de pharmacologie moléculaire et cellulaire du CNRS Sophia-Antipolis. Les sondes dessinées présentent une taille de 60-mer, un Tm compris entre 70 et 92°C, et un GC% compris entre 20 et 60%.

Les nouvelles puces dessinées présentent un total de 16 116 sondes spécifiques d'*E. ruminantium* et de 83 sondes bovines spottées en triplicats. Ces sondes bovines, ont été incluses de manière à vérifier la contamination des échantillons en ADNc eucaryotes.

3.3. Marquage, hybridation et analyse des puces à ADN

Trois réplicats biologiques pour chacune des 4 souches étudiées (Gardel et Sénégal virulentes et atténuées) ont été produits. Nous avons hybridé chacun des réplicats aux trois temps post-infections correspondant à trois stades de développement différents (T1, T2, T3). Le temps T1, étant le temps précoce pour lequel nous avons très peu de bactéries, nous avons hybridé uniquement les échantillons capturés 2 fois par SCOTS (T1 2X). Pour le temps intermédiaire T2 et le temps de lyse T3, nous avons hybridé à

la fois les échantillons capturés 2 fois par SCOTS (T2 2X et T3 2X) mais aussi, les échantillons sans capture SCOTS (T2 0X et T3 0X).

Le protocole de marquage et d'hybridation suivi est le protocole Agilent « Oligonucleotide Array-Based CGH for Genomic DNA Analysis » version 7.1. Ce protocole utilisé pour le marquage et l'hybridation d'ADN génomique a été adapté à la nature de nos échantillons ADNc. Les échantillons ont été tous marqués au Cy3. La lecture de la fluorescence après hybridation est réalisée à l'aide du scanner Agilent C, puis l'acquisition et l'analyse de l'image, à l'aide du logiciel « Feature Extraction software version 10.10 ». Les images en format .tif ainsi que les fichiers de quantifications sont ensuite stockés sur le système d'information Mediante ([Le Brigand & Barbry, 2007](#)) permettant de récupérer facilement les données pour l'analyse statistique.

Le marquage, l'hybridation, ainsi que l'acquisition et l'analyse des puces à ADN ont été effectués sur la plate-forme de génomique fonctionnelle du CNRS de Sophia-Antipolis.

3.4. Analyses statistiques des résultats

Pour l'analyse statistiques, deux approches sont utilisées : la première à l'aide du package d'analyse limmaGUI disponible sous le logiciel "R" (<http://cran.rproject.org/>) une normalisation inter-lame a été effectuée par la méthode quantile qui permet d'homogénéiser la distribution des intensités entre les lames, l'analyse statistique des données est ensuite effectuée et permet d'obtenir une liste de gènes différentiellement exprimés à l'aide d'un score statistique dérivé d'un modèle linéaire utilisant une approche empirique bayésienne ; la seconde est basée sur l'analyse de variance effectuée sur les valeurs d'intensité normalisées (quantile) puis transformées en *logarithme* de base 2 en utilisant le logiciel GeneANOVA ([Didier, et al., 2002](#)).

Pour l'analyse du transcriptome, les valeurs d'intensité obtenues ont été moyennées entre les réplicats et les valeurs de fold change pour chaque sonde ont été calculées sur ces intensités moyennes. Les sondes qui ne s'allument pas ont été éliminées.

Un gène est considéré différentiellement exprimé si 60% des sondes spécifiques de ce gène le sont.

3.5. Comparaison possibles des résultats

Nous avons généré les résultats microarrays avant et après capture pour les 4 souches : Gardel virulente et atténuée, Sénégal virulente et atténuée, aux trois stades de développement T1, T2 et T3.

La première comparaison possible est une comparaison à chaque temps post-infection du niveau d'expression des gènes entre souches virulente et atténuée (flèches bleues, figure 8). La liste des gènes différentiellement exprimés entre souche virulente et atténuée pour Gardel peut être confrontée à celle obtenue pour Sénégal. Nous avons alors des gènes communs aux 2 souches ou spécifiques de Gardel et Sénégal. En comparant l'expression des gènes pour un temps donné entre les échantillons non capturés et capturés par SCOTS (flèches rouges, figure 8), on peut mesurer l'impact de la capture sur l'expression et identifier ainsi des gènes faiblement exprimés.

La deuxième comparaison possible est une comparaison du niveau d'expression des gènes au cours du temps pour une souche donnée, virulente ou atténuée (flèches jaunes, figure 8). L'expression des gènes peut être comparée aux temps T1, T2 et T3 sur des échantillons capturés (SCOTS 2X) et aux temps T2 et T3 sur des échantillons sans captures. Ces comparaisons permettent d'identifier les gènes différentiellement exprimés au cours du développement de la souche considérée.

Dans le cadre de ma thèse, la comparaison de l'expression des gènes a été réalisée:

- au cours du développement (T2 *vs* T3) pour la souche Gardel virulente sur des échantillons capturés 2 fois
- pour le temps T3 entre souche virulente et atténuée (Gardel et Sénégal) sur des échantillons non capturés (SCOTS 0X).

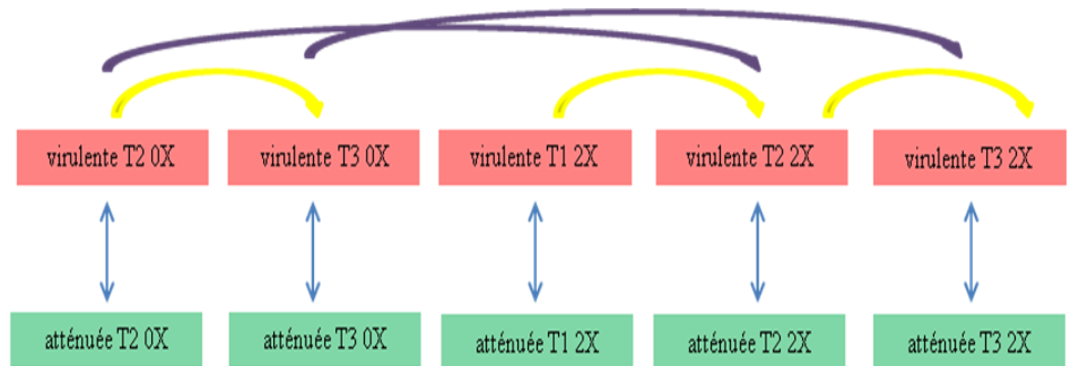


Figure 8 : Schéma récapitulatif des différentes comparaisons possible

3.6. Annotation des gènes différentiellement exprimés

Les sondes ont été identifiées selon l'annotation des génomes de Gardel virulente et Welgevonden Erwe et Erwo permettant ainsi d'identifier les CDS correspondants. Nous avons ainsi généré la liste des sondes associée aux gènes différentiellement exprimés. La fonction des gènes a ensuite été identifiée grâce à la base de données NCBI (Nucleotide database). Nous avons ensuite regroupé ces gènes en fonction de leur COG (Clusters of Orthologous Genes). Pour les gènes hypothétiques, la banque de données InterProt (Protein sequence analysis & classification) a été utilisée pour identifier des domaines protéiques connus, permettant ainsi l'assignation de certains gènes hypothétiques à des fonctions connues.

Pour chacun des gènes identifiés comme différentiellement exprimés, une analyse des séquences entre virulentes et atténuées a été effectuée, afin de vérifier la présence de mutations éventuelles pouvant altérer l'expression du gène.

3.7. Validation de certains gènes par q-RTPCR

Une validation par q-RTPCR a été effectuée pour certains gènes différentiellement exprimés entre les stades de développement chez la souche Gardel virulente. Ces gènes ont été mis en évidence par microarrays sur les échantillons capturés deux fois (SCOTS 2X). La validation du différentiel d'expression de ces gènes, a été effectuée sur 4 replicats biologiques de la souche Gardel virulente.

Deux kits de conversion en cDNA ont été utilisés. Il s'agit du kit *First Strand cDNA Synthesis kit* (GE Healthcare) et du kit *SuperScript Vilo cDNA synthesis kit* (Invitrogen).

La mesure de l'expression des gènes d'intérêt a été effectuée en q-PCR Sybergreen avec des amorces spécifiques de chaque gène. La liste des amorces, ainsi que les conditions de PCR sont fournies dans l'article correspondant ([Pruneau, et al., 2012](#)).

Afin de normaliser l'expression des gènes et ne possédant pas de gènes de ménage stables d'*E. ruminantium* pouvant servir de normalisateur (quantification relative), le nombre de bactéries par échantillon a été quantifié par q-PCR et rapporté au nombre de transcrits mesuré (quantification absolue) :

$$R_T = (\text{nombre de copies d'ADNc}) / (\text{nombre de copies d'ADNg})$$

L'expression différentielle entre les deux stades de développement était ensuite calculée à l'aide de la formule suivante :

$$R = R_{Tx} / R_{Ty}$$

Les ratios ont été convertis en log de 2 et les résultats ont été exprimés en « nombre de fois surexprimé » (fold change-FC) entre les deux stades de développement.

III. Chapitre 3 : Résultats

SOMMAIRE

1. Réplicats biologiques Gardel et Sénégal produits	102
1.1. Développement <i>in vitro</i> des souches étudiées.....	102
1.2. Quantification bactérienne des réplicats par qPCR.....	104
1.3. Choix des temps post-infection pour l'analyse du transcriptome des 4 souches	108
1.4. Vérification des captures SCOTS par q-PCR ciblant l'ARN 16S.....	109
2. Analyse du transcriptome des souches Gardel et Sénégal virulente et atténuée	111
2.1. Vérification de l'homogénéité des réplicats	111
2.2. Comparaison globale des intensités obtenus entre les échantillons sans captures (T2 0X et T3 0X) et les échantillons capturés par SCOTS (T2 2X et T3 2X).	114
2.3. Analyse du transcriptome de la souche Gardel virulente au cours du développement -Global gene expression profiling of <i>Ehrlichia ruminantium</i> at different stages of development, FEMS Immunology and Medical Microbiology, 2011	117
2.4. Analyse du transcriptome des souches Gardel et Sénégal virulente et atténuée au stade corps élémentaire (T3). Transcriptome of virulent and attenuated strains of <i>Ehrlichia ruminantium</i> reveal important genes in attenuation mechanisms (article en préparation).	128

1. Réplicats biologiques Gardel et Sénégal produits

Pour chaque souche, nous avons produit 3 réplicats biologiques différents. Il s'agit des réplicats Gp38, Gp43 et Gp44 pour Gardel virulente ; Gp238, Gp243 et Gp251 pour Gardel atténuée, les nombres indiquant les passages successifs sur cellules endothéliales. Concernant la souche Sénégal virulente les réplicats produits sont les suivants : Sp7, Sp8a et Sp8b et pour Sénégal atténuée : Sp70, Sp75 et Sp77.

Pour la validation par q-RTPCR des gènes différentiellement exprimés au cours du développement de la souche Gardel virulente, deux réplicats supplémentaires ont été produits, il s'agit des réplicats Gp41 et Gp52.

1.1. Développement *in vitro* des souches étudiées

Les 4 souches étudiées présentent des différences majeures dans leur développement *in vitro*. La première différence observée, concerne le temps du cycle de développement qui s'achève par la lyse des cellules hôtes eucaryotes. Gardel et Sénégal atténuées ont une lyse complète au bout de 4 jours tandis que Gardel et Sénégal virulentes lysent au bout de 5 et 6 jours respectivement.

Les colorations ont permis de suivre l'évolution de l'infection en temps réel. Différentes phases du cycle de développement ont pu être observées notamment l'apparition des morulas et la lyse cellulaire. Les observations réalisées sont les suivantes :

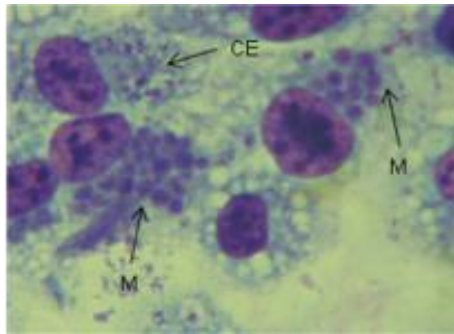
- Généralement à 24 hpi, le tapis cellulaire présente quelques gonflements signes d'une bonne infection. Il n'y a pas d'effets cytotoxiques observés et pas de différences notables entre les souches Gardel et Sénégal.
- A 48 hpi, on observe l'apparition de condensations en forme de strie sur l'ensemble du tapis cellulaire avec parfois l'observation de morulas pour la souche Gardel atténuée. Pour les souches Sénégal virulente et atténuée et Gardel virulente, il n'y a pas de différences avec l'observation à 24 hpi.

- A 72 hpi, la souche Gardel atténuée présente des morulas plus grosses et en grande quantité alors qu'on commence à détecter les morulas chez Gardel virulent. Chez Sénégal atténuée, on détecte également des morulas mais celles-ci sont beaucoup plus petites que celles de Gardel et sont organisées en grappe de raisins (Figure 9). Pour Sénégal virulente, il n'y a pas de différences avec les observations à 24 et 48 hpi.

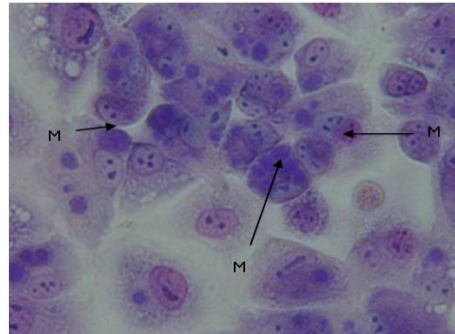
- A 96 hpi, de grosses morulas sont visibles chez Gardel virulent et il y a très peu de lyse (<10% du tapis cellulaire). Cependant les cellules sont extrêmement condensées (90% du tapis cellulaire). Pour la souche Sénégal virulente, on commence à détecter des morulas à l'intérieur des cellules eucaryotes. Pour les souches Gardel et Sénégal atténuées, les cellules sont complètement lysées, avec une lyse proche de 100%. La coloration permet de mettre en évidence de nombreuses morulas par cellule et on observe dans le milieu de culture des CE au niveau des points d'éclatement des morulas.

- A 120 hpi, une lyse supérieure à 80% est observé pour Gardel virulente. Pour Sénégal virulente, on observe une très forte condensation des cellules ainsi que la présence de nombreuses morulas à l'intérieur de celles-ci, mais il y a très peu de lyse observée.

-A 144 hpi, la lyse totale des cellules est observée pour la souche Sénégal virulente.



Sénégal virulente



Gardel virulente

Figure 9: Différences morphologiques des morulas des souches d'*E. ruminantium* Sénégal et Gardel observées par microscopie optique dans des cellules endothéliales aortiques bovines (BAE) infectées *in vitro*. Cliché du CIRAD.

1.2. Quantification bactérienne des réplicats par qPCR

a) Gardel virulente

Les courbes de croissance obtenues pour la souche Gardel virulente, ont un profil classique de croissance bactérienne avec une phase de latence entre 24 et 48 hpi, une phase de croissance exponentielle entre 48 et 96 hpi et le plateau qui est atteint entre 96 hpi et 120 hpi (Figure 10). Les trois réplicats biologiques présentaient des profils similaires avec des quantités équivalentes aux différents temps post-infection. Nous pouvons considérer qu'ici, nous avons une bonne synchronisation des réplicats biologiques.

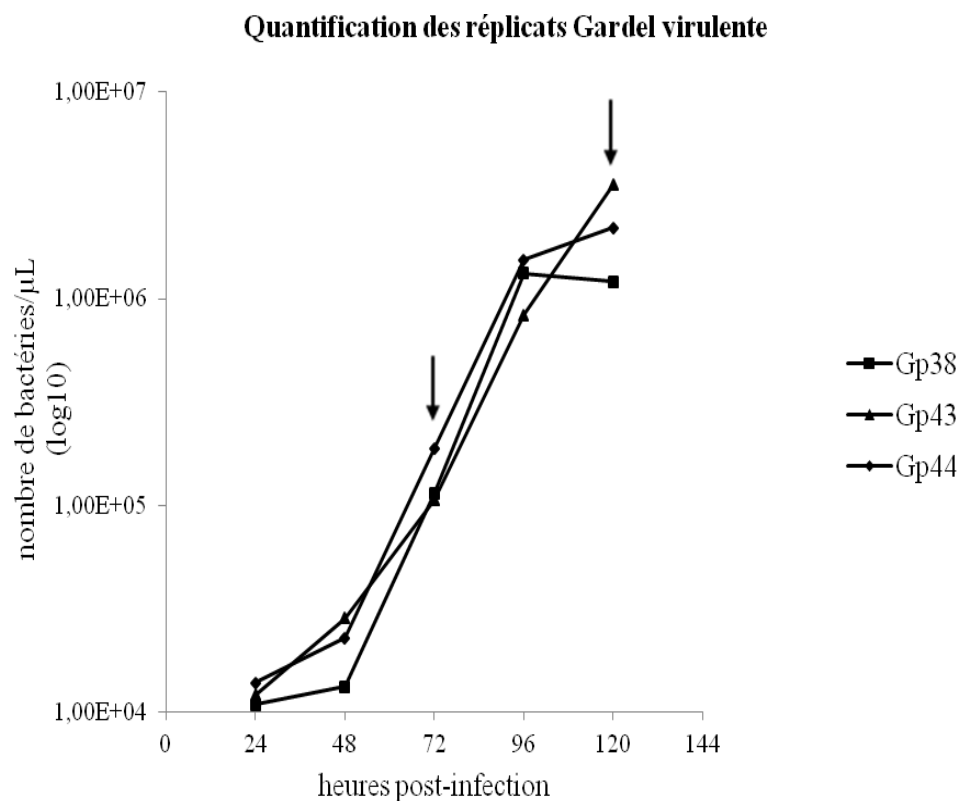


Figure 10 : Courbes de croissance des réplicats Gardel virulente

b) Gardel atténuée

Pour Gardel p238, la courbe de croissance de la souche Gardel atténuée démarre directement par la phase exponentielle de croissance à 24 hpi avec une quantité jusqu'à 10 fois plus importante que pour la souche Gardel virulente, puis entre 72 hpi et 96 hpi on atteint la phase plateau (Figure 11). Pour les 2 autres courbes de croissance, il n'y a pas de phase plateau entre 72 et 96 hpi.

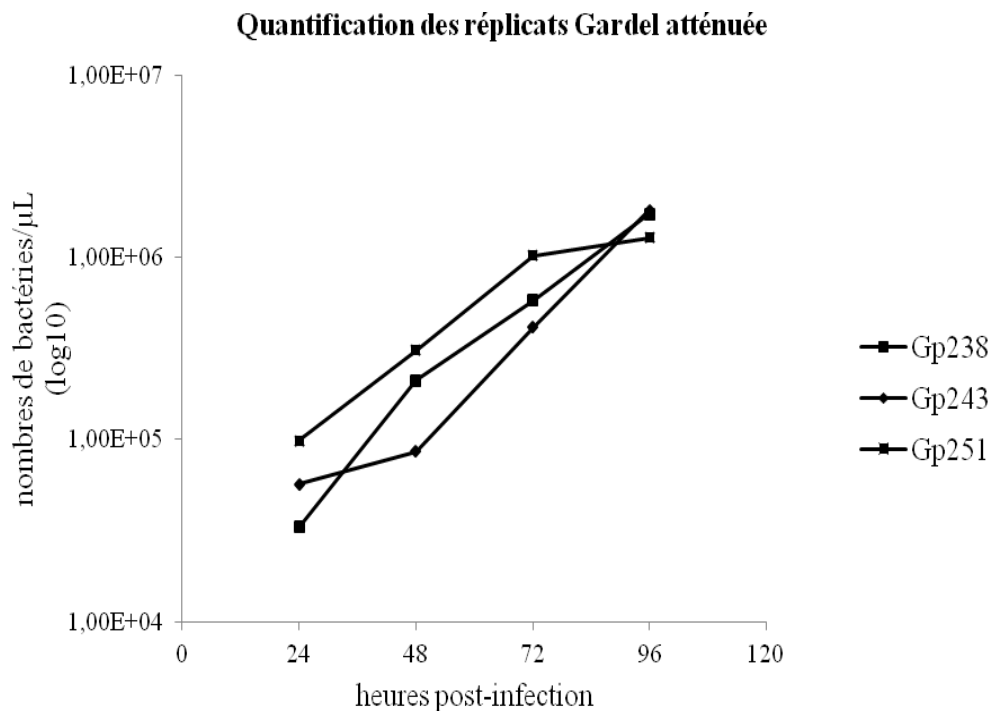


Figure 11 : Courbes de croissance des réplicats Gardel atténuée

c) Sénégal virulente

Les courbes de croissance obtenues pour les réplicats de la souche Sénégal virulente, présentent un profil différent les uns des autres. En se référant aux profils de cinétique obtenus pour les réplicats Sp8a et Sp8b, nous pouvons distinguer trois phases de croissance : une longue phase de latence entre 24 hpi et 72 hpi, une phase exponentielle de croissance entre 72 hpi et 120 hpi et une phase plateau entre 120 hpi et 144 hpi (Figure 12). Pour Sp7b, la phase de latence semble plus courte que pour les 2 autres.

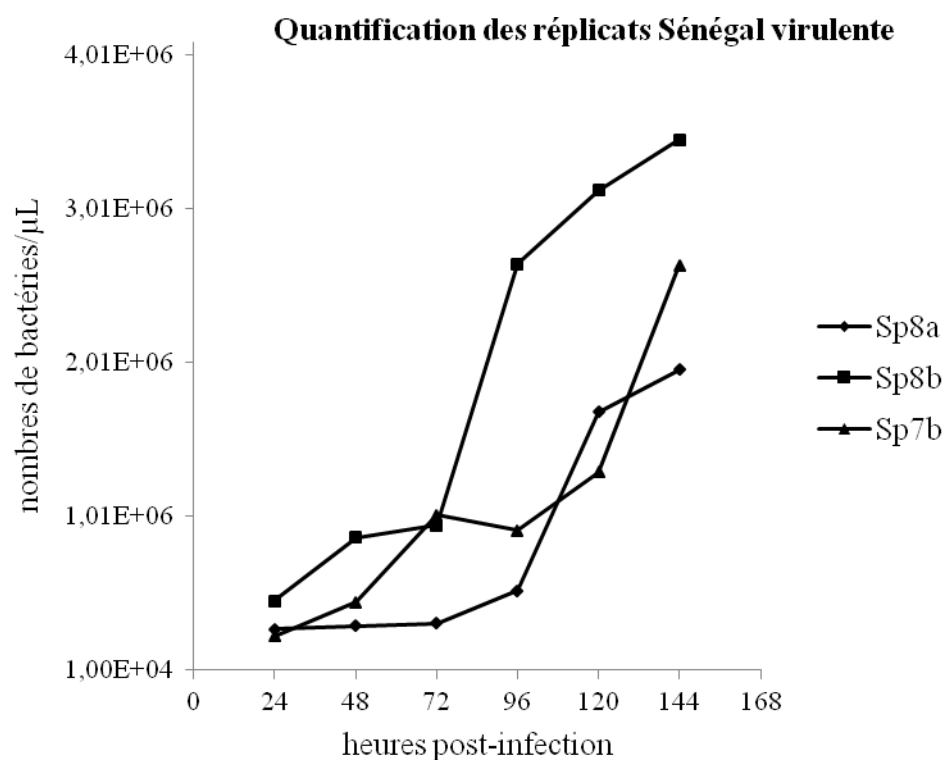


Figure 12 : Courbes de croissance des réplicats Sénégal virulente

d) Sénégal atténuée

La souche Sénégal atténuée lyse au bout de 96 hpi comme la souche Gardel atténuée, mais avec une courbe de croissance différente de celle de Gardel atténuée. En effet, Sénégal atténuée présente une phase de latence entre 24 hpi et 48 hpi, suivi de la phase exponentielle de croissance de 48 hpi à 96 hpi (Figure 13). Les quantités de bactéries aux différents temps-post infection sont très variables entre les réplicats biologiques mais elles suivent globalement le même profil de croissance.

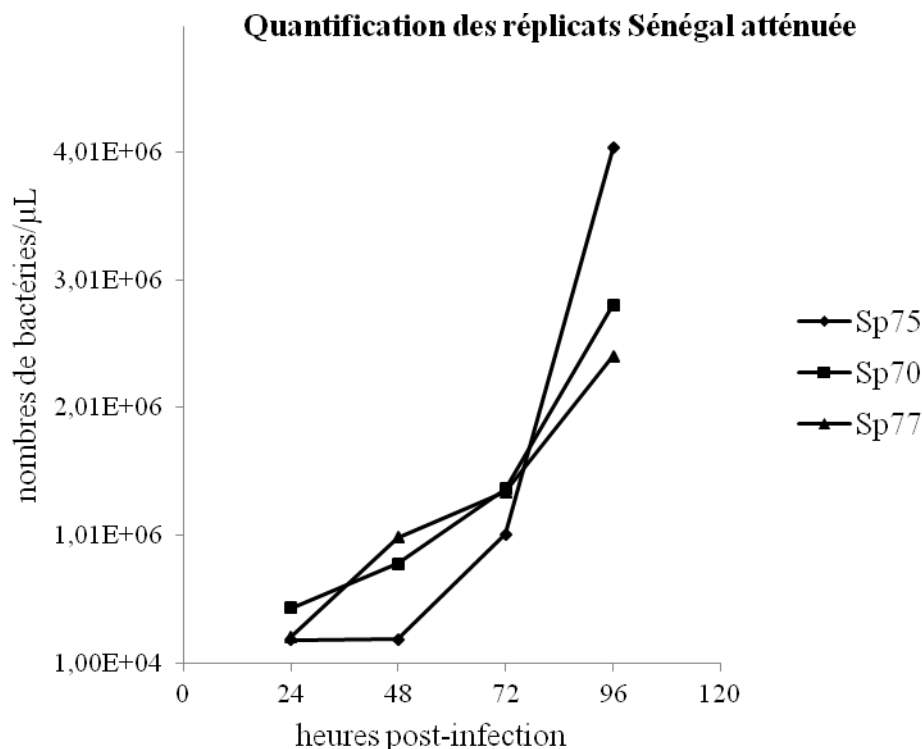


Figure 13 : Courbes de croissance des répliquats Sénégal atténuée

1.3. Choix des temps post-infection pour l'analyse du transcriptome des 4 souches

La quantification du nombre de bactéries aux cours du temps et l'observation des différentes phases évolutives d'*E. ruminantium* au microscope tout au long de l'infection ont permis de choisir les 3 temps post-infection (T1, T2, T3) pour lesquels l'expression des gènes sera comparée entre les souches Gardel et Sénégal virulente et atténuée. Selon les souches, les temps post infection correspondant à un stade donné de développement sont différents. D'après les résultats de la quantification bactérienne à 24 hpi, les deux souches Gardel (figures 10 et 11) se situent pendant ou à la fin de la phase de latence (Gardel atténuée), ce que l'on identifiera comme le temps de comparaison 1 (T1). Pour Sénégal, qui globalement présente une quantité moindre de bactéries, nous avons fait le

choix de prendre pour le temps T1, la fin de la phase de latence qui correspond à 48 hpi pour Sénégal atténuée et 72 hpi pour Sénégal virulente (Tableau 2). Le temps T2 correspond à la phase exponentielle de croissance. Ce temps a été déterminé à 72 hpi pour les souches Gardel virulente et Sénégal atténuée, à 48 hpi pour la souche Gardel atténuée, et à 96 hpi pour la souche Sénégal virulente. Enfin, le temps T3 qui est le temps de la lyse complète des cellules hôtes infectées, correspond à 96 hpi pour les souches Gardel et Sénégal atténuées, 120 hpi pour la souche Gardel virulente et 144 hpi pour la souche Sénégal virulente. A ce stade, les bactéries sont toutes présentes sous la forme libre infectieuse.

	Gardel virulente	Gardel atténuée	Sénégal virulente	Sénégal atténuée	Stade	Observations au microscope
T1	24 hpi	24 hpi	72 hpi	48 hpi	Phase de latence	Condensation
T2	72 hpi	48 hpi	96 hpi	72 hpi	Croissance exponentielle	Petites morulas
T3	120 hpi	96 hpi	144 hpi	96 hpi	Lyse cellulaire	Grosses morulas & CE

Tableau 2 : Temps T1, T2, et T3 sélectionnés pour chacune des souches et stades de croissance correspondant

1.4. Vérification des captures SCOTS par q-PCR ciblant l'ARN 16S

La méthode sélective de capture de transcrits (SCOTS) permet l'élimination de l'ARN eucaryote et la diminution des transcrits ARN ribosomaux procaryotes qui représentent 80% de l'ARN total procaryote. Après captures SCOTS de nos différents échantillons, nous avons vérifié cette diminution en ciblant l'ADNc issu de la conversion de l'ARN 16S, cette diminution est associée à l'efficacité de la méthode de capture. Nous avons

donc quantifié pour chaque temps (T1, T2, T3) et pour chaque round de capture (0X, 1X, 2X), le nombre de transcrits 16S par q-PCR. Les résultats obtenus pour les répliquats Gardel atténuée p243 et Sénégal virulente p8a sont présentés dans la figure 14. Il y a une diminution au moins de moitié de la quantité de transcrit 16S au temps T3 après la première capture pour Gardel et Sénégal. Au temps T2, l'effet est marqué uniquement pour Sénégal. Curieusement au temps T1, il y a une augmentation de la quantité de transcrit pour Sénégal au cours des captures.

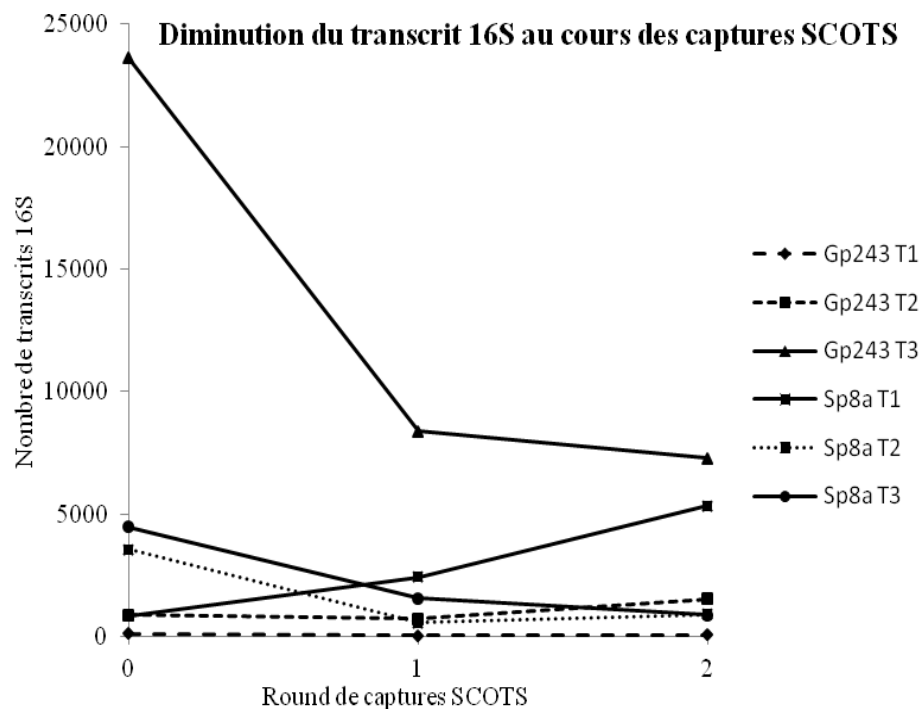


Figure 14: Quantification du transcrit 16S au cours des captures pour les répliquats Gp243 et Sp8a.

2. Analyse du transcriptome des souches Gardel et Sénégal virulente et atténuée

2.1. Vérification de l'homogénéité des réplicats

La première étape de notre analyse a été de vérifier pour chaque condition expérimentale, l'homogénéité des résultats obtenus entre les différents réplicats. Pour cela, nous avons calculé le coefficient de corrélation linéaire entre les valeurs d'intensités de fluorescence transformées en log2 des réplicats pour une condition donnée. Pour une condition expérimentale, les intensités des signaux obtenues pour deux réplicats sont considérées comme étant homogènes si le coefficient de corrélation linéaire calculé est supérieur à 0,6.

Les résultats des coefficients de corrélation obtenus sont présentés par souches: Gardel virulente (figure 15), Gardel atténuée (figure 16), Sénégal virulente (figure 17) et Sénégal atténuée (figure 18). Pour toutes les souches mise à part Sénégal virulente et pour toutes les conditions expérimentales, les coefficients de corrélation obtenus sont supérieurs à 0,6. Les réplicats de Sénégal atténuée présentent les meilleurs résultats avec des coefficients de corrélation linéaire autour de 0,9 pour les conditions T2 0X, T3 0X et T2 2X (figure 17).

Pour Sénégal virulente, pour la condition T2 0X, aucuns des coefficients calculés n'étaient supérieur à 0,6. Ce résultat peut s'expliquer d'une part par les différences de quantité bactérienne que nous avons pu constater par l'observation des courbes de croissance de Sénégal virulente (figure 12). En effet, pour le temps de comparaison T2, il y avait une différence du nombre de bactérie de 2E+06 entre les 3 réplicats. Autre fait pouvant expliquer globalement les coefficients de corrélation plus faible pour Sénégal virulente est l'allure de la courbe de croissance du réplicat Sp7 qui est très différente des autres. En effet contrairement aux deux autres réplicats, Sp7 semble présenter une phase de latence de 24 hpi à 96 hpi.

Globalement, les différents réplicats biologiques montrent une assez bonne homogénéité entre elles.

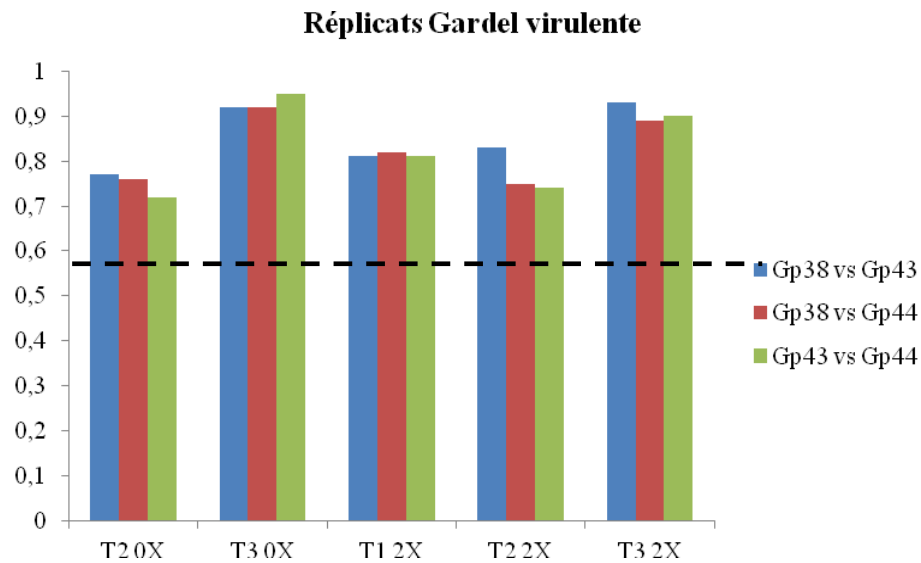


Figure 15 : Coefficients de corrélation linéaires des réplicats Gardel virulente en fonction des différentes conditions expérimentales

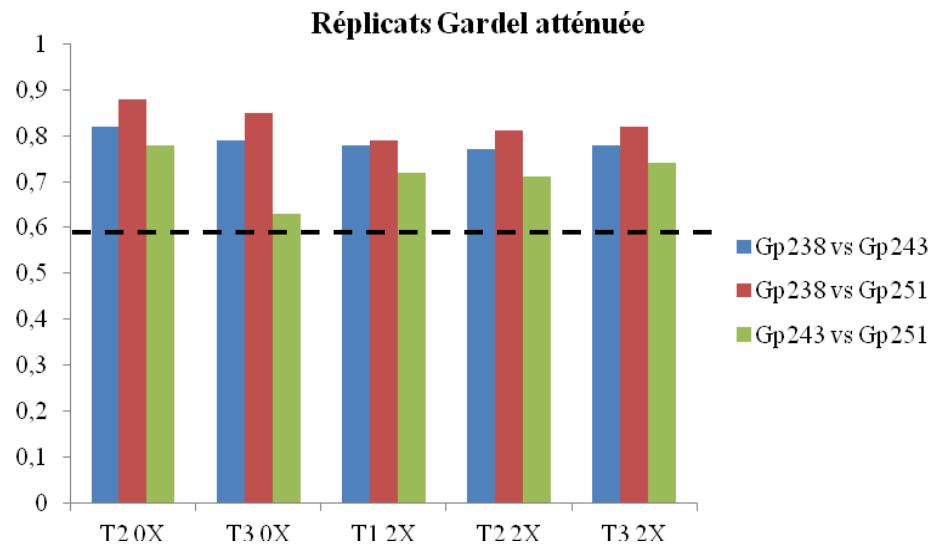


Figure 16 : Coefficients de corrélation linéaires des réplicats Gardel atténuée en fonction des différentes conditions expérimentales

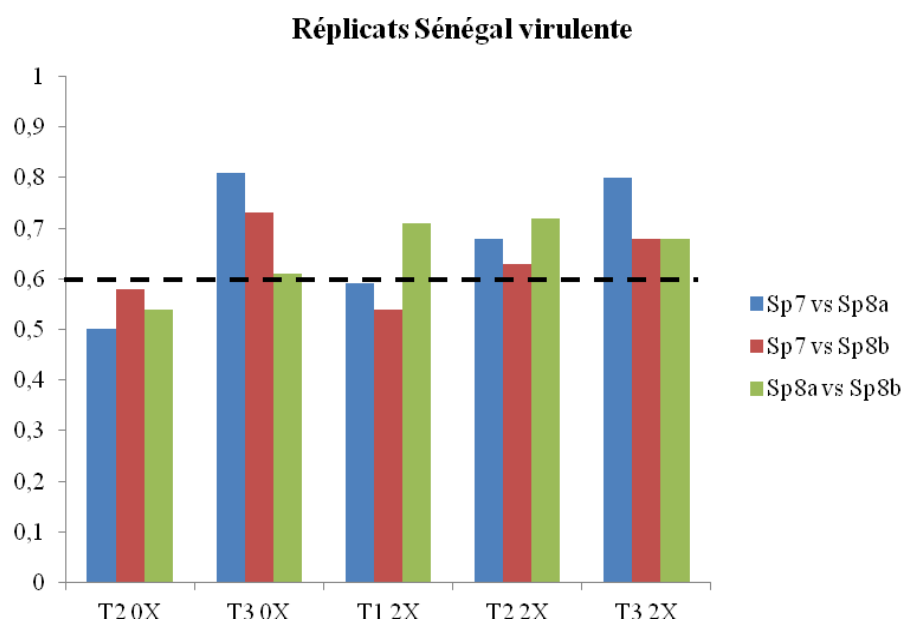


Figure 17 : Coefficients de corrélation linéaires des répliants Sénégal virulente en fonction des différentes conditions expérimentales

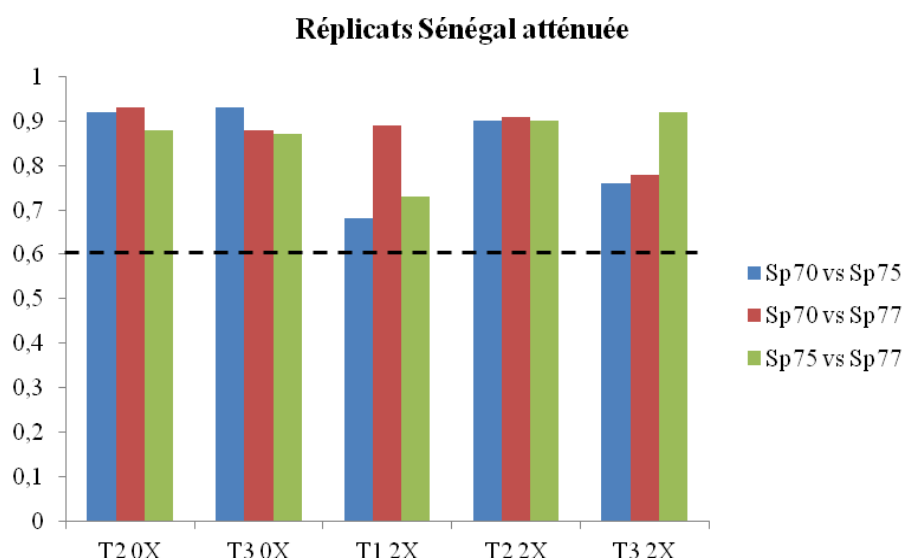


Figure 18 : Coefficients de corrélation linéaires des répliants Sénégal atténuée en fonction des différentes conditions expérimentales

2.2. Comparaison globale des intensités obtenus entre les échantillons sans captures (T2 0X et T3 0X) et les échantillons capturés par SCOTS (T2 2X et T3 2X).

Afin de pouvoir évaluer l'effet de la méthode SCOTS sur l'analyse du transcriptome d'*E. ruminantium*, nous avons comparé les résultats d'hybridation obtenus pour les échantillons sans captures avec ceux obtenus pour les échantillons capturés pour les temps T2 et T3. Dans les figures 19 et 20, ce sont les FC entre souche virulente et atténuée qui sont comparés. Les photographies des lames pour Gp238 à T2 et T3 0X et 2X sont présentées en figure 21.

Les coefficients de corrélation obtenues pour le temps T2 sont de 0,21 et 0,32 respectivement pour Gardel et Sénégal. Il n'est pas surprenant d'obtenir des coefficients de l'ordre de 0,3. En effet, nous avons déjà constaté par l'analyse des images une différence notable d'une part du nombre de sondes allumées et d'autre part, de l'intensité de fluorescence des sondes allumées. Pour le temps intermédiaire T2, la méthode SCOTS permet un enrichissement et donc une meilleure détection des transcrits d'*E. ruminantium*. Comme nous pouvons le constater sur la figure 21, nous avons une augmentation des signaux donc du nombre de sondes allumées à T2 2X comparé à T2 0X.

Pour le temps tardif T3, les coefficients de corrélation obtenues étaient de 0,43 pour Gardel et 0,06 pour Sénégal. La très forte différence observée pour Sénégal est surprenante. Des analyses complémentaires plus fine devront être effectuées afin d'identifier les biais entraînés par la méthode SCOTS pour l'analyse du transcriptome des souches Sénégal virulente et atténuée.

Cependant, l'analyse des images des différents réplicats Sénégal et Gardel au temps T3 et sans capture SCOTS, montre un pourcentage d'allumage des sondes important de 74% (12 000 / 16 116 sondes) dont 18% (3 000) avec une forte intensité. Il semblerait donc qu'au temps T3, la méthode SCOTS ne soit pas nécessaire pour la détection des transcrits exprimés. Comme nous

pouvons le constater sur la figure 21, il y'a peu de différences d'allumage des sondes entre T3 0X et T3 2X.

Comparaison des FC pour T2: SCOTS 0X vs 2X

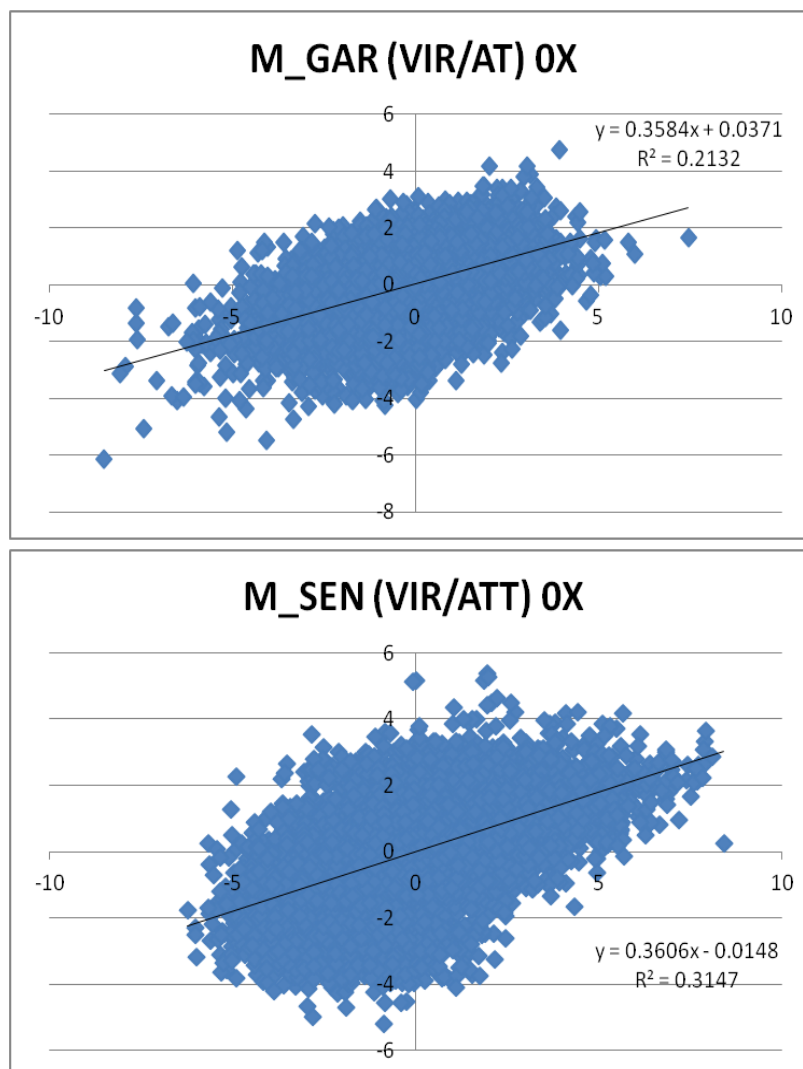


Figure 19 : Comparaison des FC pour le temps T2 entre 0X (abscisse) et 2X (ordonné) pour Gardel et Sénégal.

Comparaison des FC pour T3: SCOTS 0X vs 2X

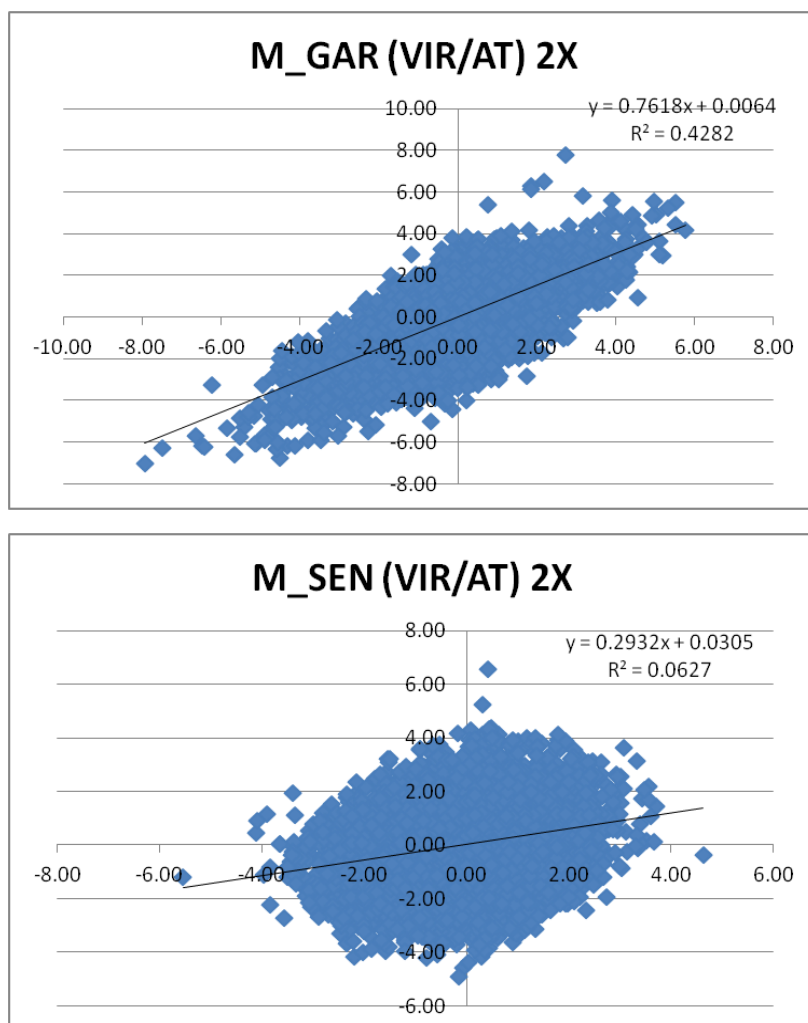


Figure 20 : Comparaison des FC pour le temps T3 entre 0X (abscisse) et 2X (ordonné) pour Gardel et Sénégal.

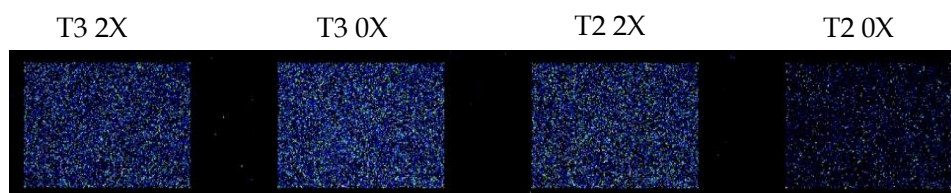


Figure 21 : Photographie des lames de Gp238 montrant les intensités de fluorescence pour les temps T2 et T3 entre 0X et 2X.

2.3. Analyse du transcriptome de la souche Gardel virulente au cours du développement -Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development, FEMS Immunology and Medical Microbiology, 2011

Afin d'identifier les mécanismes intervenant dans le développement intracellulaire d'*E. ruminantium*, nous avons analysé et comparé l'ensemble des transcrits générés par la souche Gardel virulente à 2 temps post-infection correspondant au stade intracellulaire non infectieux (corps réticulé, CR) et au stade extracellulaire infectieux (corps élémentaire, CE). Cette analyse a été effectuée à partir des données générées avec les premières puces à ADN dessinées au cours de la précédente thèse. Les gènes différentiellement exprimés aux deux stades de développement (induits ou réprimés) ont pu être identifiés. La fonction de ces gènes a ensuite été déterminée par génomique comparée lorsque la fonction était identifiée pour d'autres pathogènes, ou par la recherche de motifs protéiques connus, pour les gènes de fonctions inconnues.

La comparaison de l'expression différentielle des gènes entre les stades T1 et T2 n'a pas été effectuée car, au stade T1 il n'y avait pas suffisamment de gènes détectés pour que l'analyse soit valide. En effet,

la quantité d'ADNc contenu dans les échantillons était faible malgré l'enrichissement des transcrits par la méthode SCOTS. Pour ce temps, seulement 24% des sondes donnaient un signal d'allumage.

Cinq pour cent (54/950) des CDS prédits étaient différentiellement exprimés entre les stades CR et CE. Dix neuf gènes étaient trouvés sur-exprimés au stade CR, tandis qu'au stade CE, 35 gènes étaient sur-exprimés. A partir des différentiels observés, nous avons obtenu un véritable profil d'expression global pour *E. ruminantium*, avec l'identification de nombreux gènes ou systèmes de régulation impliqués dans le développement de la bactérie. En effet, des gènes impliqués dans les fonctions du métabolisme comme la production d'énergie, le transport et le métabolisme de coenzyme, d'acides aminés, d'ions, de nucléotides, étaient trouvés pour les deux stades de développement avec une proportion généralement plus importante pour le stade CR. Certains groupes fonctionnels de gènes étaient cependant, spécifiques d'un stade de développement. Les mécanismes de transductions des signaux et les mécanismes de défense étaient sur-exprimés exclusivement pour le stade CE, tandis que les gènes impliqués dans la biosynthèse et transport des métabolites secondaires, dans le trafic intracellulaire et la sécrétion, et les gènes codant pour la famille multi-génique *map1*, étaient sur-exprimés pour le stade CR.

Dans un second temps, sur 3 réplicats biologiques Gardel virulente, une validation du différentiel d'expression obtenu par microarrays a été effectuée par qRTPCR pour 11 gènes d'intérêt sélectionnés soit pour leur différentiels d'expression élevés ou pour leur rôle d'intérêt déterminés chez d'autres pathogènes.

Sur ces 11 gènes, le différentiel d'expression observé en microarrays était confirmé par q-RTPCR pour 7 gènes et 4 gènes avaient un différentiel d'expression inverse de celui observé en microarrays.

Parmi ces 11 gènes, nous avons confirmé la sur-expression de *dksA*, qui code pour un facteur de transcription activant des gènes de virulence chez d'autres intracellulaires comme *Salmonella typhimurium*.

L'ensemble de ces résultats a été publié dans FEMS Immunology and Medical Microbiology en 2011.

RESEARCH ARTICLE

Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development

Ludovic Pruneau¹, Loïc Emboulé¹, Pryscellia Gely¹, Isabel Marcelino^{1,2}, Bernard Mari³, Valérie Pinarello¹, Christian Sheikboudou¹, Dominique Martinez⁴, France Daigle⁵, Thierry Lefrançois¹, Damien F. Meyer¹ & Nathalie Vachieri¹

¹CIRAD, UMR CMAEE, Guadeloupe, France; ²IBET, Oeiras, Portugal; ³UMR 6097, CNRS-Université de Nice Sophia Antipolis, Institut de Pharmacologie Moléculaire et Cellulaire, Sophia Antipolis, France; ⁴CIRAD, UMR CMAEE, Montpellier, France; and ⁵Département de Microbiologie et Immunologie, Université de Montréal, Montréal, QC, Canada

Correspondence: Nathalie Vachieri, CIRAD, UMR CMAEE, F-97170 Petit-Bourg, Guadeloupe, France. Tel.: +590 590255442; fax: +590 590940396; e-mail: nathalie.vachieri@cirad.fr

Received 31 July 2011; revised 20 October 2011; accepted 31 October 2011.
Final version published online 8 December 2011.

DOI: 10.1111/j.1574-695X.2011.00901.x

Editor: Gilbert Greub

Keywords

Ehrlichia; transcriptomic; pathogenicity; development.

Abstract

Ehrlichia ruminantium (ER), the causative agent of heartwater on ruminants, is an obligate intracellular bacterium transmitted by ticks of the genus *Amblyomma*. Previous studies have shown that early stages of development may be critical for *Ehrlichia* pathogenicity. To gain insights into the biology of intracellular ER, we determined the genome-wide transcriptional profile of ER replicating inside bovine aortic endothelial cells using DNA microarrays. At intermediate and late stages of infection (reticulate and elementary bodies, respectively), a total of 54 genes were differentially expressed. Among them, we measured by q-RT-PCR the overexpression of 11 of 14 genes. A number of genes involved in metabolism, nutrient exchange, and defense mechanisms, including those involved in resistance to oxidative stress, were significantly induced in ER reticulate bodies. This is consistent with the oxidative stress condition and nutrient starvation that seem to occur in *Ehrlichia*-containing vacuoles. During the lysis stage of development, when ER is infectious, we showed the overexpression of a transcription factor, *dksA*, which is also known to induce virulence in other pathogens such as *Salmonella typhimurium*. Our results suggest a possible role of these genes in promoting ER development and pathogenicity.

Introduction

Ehrlichia ruminantium (ER) is a α -proteobacterium in the order *Rickettsiales* that is transmitted by *Amblyomma* ticks. This bacterium is the causal agent of heartwater, a fatal disease in ruminants (Allsopp, 2010). This disease represents a serious problem for livestock productivity in endemic areas such as sub-Saharan Africa and the Caribbean threatening the American continent where indigenous competent ticks are present (Barre *et al.*, 1987).

ER is an obligate intracellular pathogen that infects the endothelium of blood vessels. ER has a complex life cycle described as a *Chlamydia*-like developmental cycle (Jongejan *et al.*, 1991). In the early stage of the cycle, elementary bodies (EB), which represent the extracellular and infectious forms of the bacterium, adhere to host target cells and then are engulfed quickly. They remain within

intracytoplasmic vacuoles, where they divide by binary fission to produce the vegetative intracellular noninfectious forms, the reticulate bodies (RB), after 2–4 days postinfection, and further intermediate bodies (IB) after 4–5 days. After 5–6 days, the disruption of host cells leads to the release of numerous infective EB, thus initiating a new infectious cycle.

From the sequencing of Gardel and Welgevonden strains, whole-genome ER microarrays were designed to analyze the ER transcriptome at different development stages (Frutos *et al.*, 2006; Emboule *et al.*, 2009). For *Rickettsia conorii*, the transcriptomic analysis highlighted the overexpressed genes involved in resistance to oxidative and osmotic stress (thioredoxin, *trxB2*, and proline–betaine transporter, *proP*), in DNA repair and recombination, and in some virulence factors (Renesto *et al.*, 2008).

The aim of this study was to understand the *ER* development and pathogenicity. Sufficient amount of *ER* RNA from noninfectious and infectious forms of *ER* was produced using bovine endothelial cells cultured *in vitro* and infected with virulent strain. First, the upregulated genes between RB and EB stages of development were identified by microarray. Then, the differential gene expression between these stages was evaluated by q-RTPCR on 14 selected genes.

Materials and methods

Production of biological samples

Production of biological samples was performed using bovine aorta endothelial cells (BAE) infected with virulent Gardel strain, as described by Emboule *et al.* (2009).

Four independent experiments were performed using Gardel strain passages p38, p39, p44, and p52. An additional experiment using Gardel p41 was carried out and used for further validation when necessary. For each experiment, *ER* was produced at early (24 hpi = T1 and 48 hpi), intermediate (72 hpi = T2 and 96 hpi), and late development stages (120 hpi = T3) (Fig. 1). At each time postinfection except at 120 hpi, the cell monolayer was harvested by trypsinization and 1/10 of sample was used for DNA extraction and 9/10 for RNA extraction. When 80% cell lysis was observed, at 120 hpi, both supernatant and cellular debris were harvested and divided as described earlier. The number of bacteria per sample was quantified for each time postinfection. For gene expression measurements, only samples collected at 72 hpi, T2 and 120 hpi, T3 were used. The experiment with Gardel p39 was only used for microarray analysis. Selective capture of *ER* transcripts was carried out before hybridization on microarrays as described below. The four other experiments were used for measurements by q-RTPCR.

Quantification of *ER* in biological samples by *map-1* q-PCR

At each time postinfection, 1/10 of sample was centrifuged at 20 000 *g* for 5 min. The pellet was dissolved in 200 μ L of PBS 1 \times . Genomic DNA (gDNA) was extracted using QIAamp DNA Mini Kit (QIAGEN, France) according to the manufacturer's instructions.

The number of *ER* per sample was quantified by q-PCR on extracted DNA, at different hours postinfection. The targeted gene, *map1* present as a single copy in the *ER* genome, codes for a major antigenic protein. The primer sequences and hybridization temperature are shown in Table 1. In the PCR mix, the final concentrations of *map1* reverse and forward primers were 900 nM and the

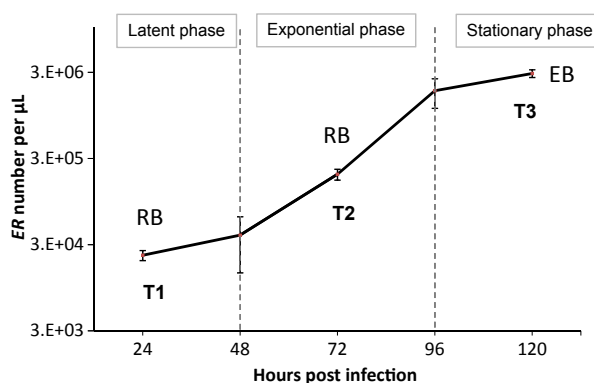


Fig. 1. Growth kinetics of *ER* Gardel strain inside bovine aortic endothelial cells as determined by quantitative PCR assay using *map-1* gene. Specific *ER* concentrations (expressed as number of *ER* per μ L) are shown as mean \pm SD of number of *ER* per μ L obtained for Gardel p38, 41, 44, and 52. The different times postinfection indicated as T1, T2, and T3 correspond to early, intermediate, and late developmental stages of *ER*, respectively. Latent, exponential, and stationary phases of *ER* growth are indicated.

final probe concentration was 250 nM. The Taqman master mix (Applied Biosystems, France) was used following the manufacturer's instructions. Four microliters of sample DNA or standard gDNA was added to the mix. The PCR program was as follows: 2 min at 50 $^{\circ}$ C, 10 min at 95 $^{\circ}$ C, and 40 cycles with 15 s at 95 $^{\circ}$ C and 1 min at 60 $^{\circ}$ C. To quantify the number of bacteria μ L $^{-1}$, a standard curve was established using gDNA of Gardel serially diluted (from 7×10^6 to 7×10^1 copies μ L $^{-1}$).

Extraction of total RNA

For each time postinfection, total RNA extraction procedure was carried out on 9/10 of sample as described by Emboule *et al.* (2009). Total RNA quantification was performed by Nanodrop 2000c (Thermo Scientific). For T2 and T3, total RNA samples were pooled in RNase-free water at a final concentration of 0.5 μ g μ L $^{-1}$ for microarray and at final concentration of 0.15 μ g μ L $^{-1}$ for qRTPCR.

Reverse transcription of RNA samples

For hybridization on *ER* microarray, RNA samples were reverse-transcribed by random priming with Superscript II (Invitrogen) according to the manufacturer's instructions. The reverse transcription and PCR amplification of corresponding cDNA were carried out using *KpnI*-primers as previously described (Emboule *et al.*, 2009).

For the validation of gene expression by q-RTPCR, RNA samples were reverse-transcribed using SuperScript VILO cDNA Synthesis kit (Invitrogen) for Gardel p38

Table 1. *ER* primers and PCR conditions for q-PCR and q-RTPCR

Primer name	Primer sequence (5'→3')	Hybridization temperature (°C)	Product size (bp)	Target gene
<i>map1</i> gardF <i>map1</i> gardR <i>map1</i> gard-ex3	CACTTGAAGGAATGCCAGTTTCTC CTTAGGATTGTAGCATTGATTACTGACACT 6-FAM-ATGCCTGCACACACATAT-MGB	60	85	<i>map1</i>
F- <i>map1</i> -6 R- <i>map1</i> -6	ATACACCAACATTCCAGAACA CAGGGATTCTGCATCGA	60	120	<i>map1</i> -6
F- <i>secF</i> R- <i>secF</i>	TGGCCCAAGTAGGGTATAAGCA ACCAAGCAAGCTAATAAGACCCAATGT	60	175	<i>secF</i>
F- <i>sppA</i> R- <i>sppA</i>	TGGCGTACTTTAACATTCTTATGGTT ACCAGTAACTGCACCACCTGGACT	60	231	<i>sppA</i>
F- <i>dapA</i> R- <i>dapA</i>	AGGCATATTTGCCGGTGATGGGT TGCTCATGGTGGTTCTGGATGC	63	182	<i>dapA</i>
F- <i>dksA</i> R- <i>dksA</i>	GCTGACACAGATCTAACAGACATGGCA TGGGTTTGCTTTGAGTCGTGCT	60	189	<i>dksA</i>
F- <i>lolD</i> R- <i>lolD</i>	TGTTCAAGCTACAACCTCAGATCGAGA CGACCCGTTGCTTTCTCTCTCCA	60	250	<i>lolD</i>
F- <i>proP1</i> R- <i>proP1</i>	TGGTGGTGAAGCAGGTGCAA GGAACTCTCCACCCCAAACA	63	183	<i>proP1</i>
F- <i>trx</i> R- <i>trx</i>	TGGGCTCCGTGGTGTGGACC AGTGGTACTGCTAACACCATACTG	63	150	<i>trx</i>
F- <i>resol</i> R- <i>resol</i>	TCAATGGTAATTGGGTGCCACTTGAT TTTGCTACCCTAGTAGCCATAGCAGT	63	158	<i>resol</i>
F- <i>ccmB</i> R- <i>ccmB</i>	TGCACTGGGTTTATGTTGGAATACCTG AGCCAATGCATGACCTACAGCAGA	60	161	<i>ccmB</i>
F- <i>cytoC1</i> R- <i>cytoC1</i>	ACGTGATGTTGGGTTTTAGAGGATG AGCAGCTACAGCAGCTCCTT	60	160	<i>cyto C1</i>
F- <i>atpB</i> R- <i>atpB</i>	AGGCACACCAATGTGGTTAGCACC TGGCCAGCTATCATATTAGCAGCGAGT	60	111	<i>atpB</i>
F- <i>hypo</i> R- <i>hypo</i>	CCCAGCGTACAACAGCTAAGGC CCATGCTGTTTGTTCAGATGTTCA	63	163	<i>hypo</i>
F- <i>folK</i> R- <i>folK</i>	TGGCTTACTCACCTGGCAATGAT GCATTTTCTGGTAATAGTGCCTGCTT	63	157	<i>folK</i>

F, forward primer; R, reverse primer.

and p52 samples and with First Strand cDNA Synthesis kit (GE Healthcare) for Gardel p41 and p44. The conversion conditions were performed according to the manufacturer's instructions.

The bacterial gDNA contaminant in RNA samples was evaluated by PCR 25 cycles targeting *pCS20* gene using primers AB128 and AB129 as described previously (Martinez *et al.*, 2004). In RNA samples, no signal was obtained by *pCS20* PCR. Moreover, the efficiency of conversion was checked using q-RTPCR targeting *ER* 16S gene by process-

ing simultaneously RNA and cDNA samples (Emboule *et al.*, 2009). The difference in Ct between RNA and cDNA samples was always higher than 5 cycles, indicating low contamination by gDNA in RNA samples.

Measure of differential gene expression by microarray

Specific cDNAs of *ER* were selected using the SCOTS method as described by Emboule *et al.* (2009). Two

successive SCOTS captures were generated for the sample used for microarray analysis.

ER microarray used in this study was previously used in the study on SCOTS method validation (Emboule *et al.*, 2009). Microarray results were analyzed using GENE-ANOVA software (Didier *et al.*, 2002). Genes differentially expressed between T2 and T3 had a \log_2 -fold change (FC) > 1 with a *P*-value < 0.1 and variance > 0.5.

The functions of genes differentially expressed and of proteins were checked on NCBI and KEGG databases, respectively.

Measure of differential expression by q-RTPCR

On the genes differentially expressed identified by microarrays, a selection of genes for validation by q-RTPCR was made. The design of the primers was carried out with PRIMER3PLUS software (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) using conventional parameters (Table 1). Primers' specificity was verified by BLAST and by the use of DNA from BAE cells as negative control during PCR validation steps. The final concentration of forward and reverse primers in the PCR mix was identical and varied from 150 to 300 nM depending on the targeted gene. SyberGreen Master Mix (Applied Biosystems) was used following the manufacturer's instructions with 2 μ L of cDNA or standard gDNA. The PCR program was 10 min at 95 °C, and 40 cycles with 30 s at 95 °C, 30 s at the optimal hybridization temperature (Table 1), and 1 min at 72 °C. An absolute quantification was carried out to measure the gene expression. A range of 7×10^6 to 7×10^1 copies μ L⁻¹ of gDNA Gardel was used for standard calibration and processed simultaneously with *ER* cDNA samples.

To normalize the gene expression, the number of cDNA copies from the targeted *ER* gene is divided by the number of *ER* (number of gDNA copies) present in the biological sample, giving *R* at one time postinfection (T):

$$R_T = (\text{number of cDNA copies})/(\text{number of } ER)$$

The differential gene expression (FC) between different development stages is measured by the following formula:

$$FC = R_{T3}/R_{T2}$$

Results are expressed in \log_2 -FC between T3 and T2. Positive \log_2 -FC corresponds to overexpression at T3 compared with T2. Negative \log_2 -FC corresponds to overexpression at T2 and is expressed as absolute value of \log_2 -FC. All qPCR were performed on the ABI Prism 7500 (Applied Biosystems).

Results

Quantification of the number of bacteria by quantitative PCR using *map-1* gene

The number of *ER* per sample was measured by q-PCR using *map-1* gene. The growth curve of *ER* had a classical profile of bacterial growth with three phases: a latent phase between 24 and 48 hpi with $2.26\text{--}3.87 \times 10^4$ bacteria μ L⁻¹, an exponential phase between 48 and 96 hpi with a 2 \log_{10} increase of *ER* concentration, and a stationary phase between 96 and 120 hpi before the lysis of the cells (Fig. 1). The final mean concentration was 3.10^6 bacteria μ L⁻¹. The *ER* development was synchronized because of standardized inoculum and its development was observed by optical microscopy at each time postinfection. The stage of *ER* between 24 (T1) and 72 (T2) hpi corresponded to the RB and the 96 hpi stage represented the IB. At 120 hpi (T3), *ER* is on its infective EB form and induces host cell lysis.

Functional category classification of genes differentially expressed during *ER* cycle

Fifty-four genes (5% of *ER* predicted genes) were found differentially expressed between T2 (RB) and T3 (EB) after microarray analysis. Depending on the gene, the \log_2 -FC of overexpression varied from 1.5 to 6.1 for RB stage and from 3 to 8.2 for EB stage (Table 2). They were classified according to their COG (clusters of orthologous groups of proteins) function and by the stage of development. The proportion of overexpressed gene categories at EB stage and RB stage is shown in Fig. 2a and b, respectively. Hypothetical genes represented the highest percentage of overexpressed genes. There were 51.4% and 26.3% of hypothetical genes overexpressed at EB and RB stages, respectively. Between the two stages, there was similar percentage of overexpressed genes (5.7% vs. 5.3%) involved in nucleotide transport and metabolism. The percentage of genes involved in (1) post-translational modification, protein turnover, and chaperones and (2) carbohydrate, AA, and inorganic ion transport and metabolism was higher for EB than for RB stages. *Rne* and *rsme* belonging to the same translation, ribosomal structure and biogenesis category were overexpressed at RB and EB, respectively. For coenzyme transport and metabolism function, there was a higher proportion of overexpressed genes at RB compared with EB stage (15.8% vs. 5.7%). Similarly, the proportion of genes belonging to replication and DNA repair and energy production and conversion groups was twice higher at RB stage (10.5% vs. 5.7%). The genes of five function

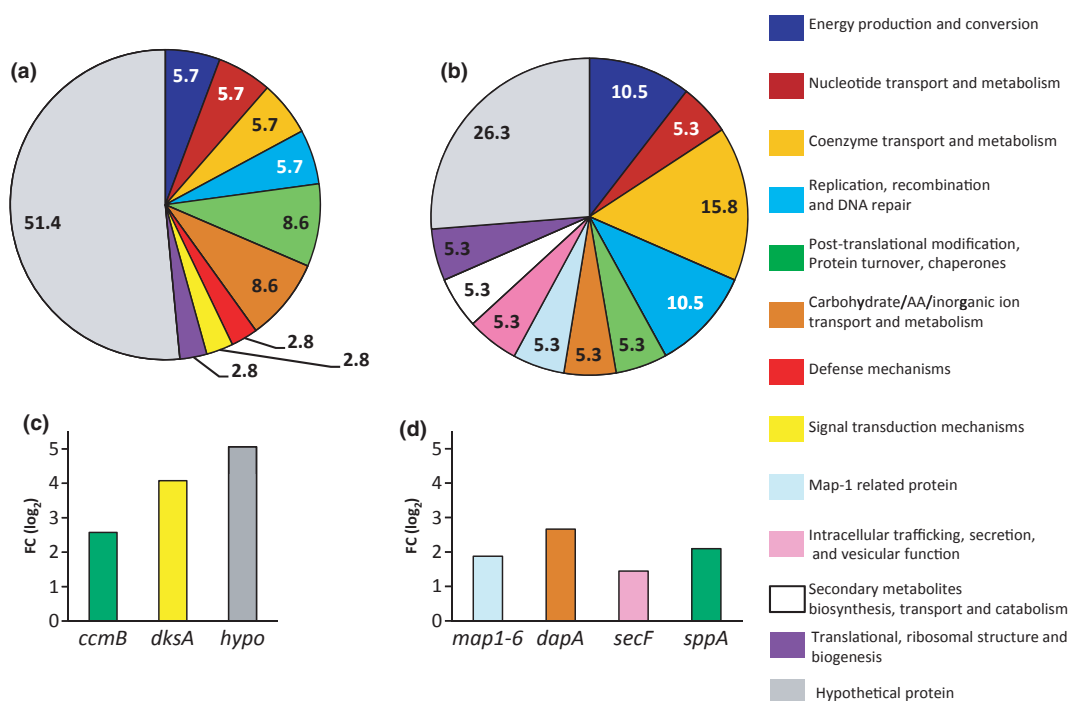


Fig. 2. Gene overexpression analysis at EB and RB stages by microarrays and q-RTPCR. Pie charts showing functional category classification of genes identified as overexpressed by microarray analysis during *ER* developmental cycle (a and b). The figure represents the proportion of genes overexpressed according to the COG functional classification for the EB (a) and RB (b) stages. Differential gene expression measured on Gardel p44 by q-RTPCR at EB (c) and RB stages (d). Experiments were repeated at least three times.

categories were specifically overexpressed at one stage. The *map1-6* gene codes for a MAP-1-related protein, and *secF* is associated with intracellular trafficking, secretion, and vesicular function at RB stage. *FabG*, involved in the secondary metabolite biosynthesis transport and catabolism, was also upregulated at RB. At EB stage, *dksA* is involved in signal transduction mechanisms and *lolD* is taking part in defense mechanisms (Fig. 2a and b and Table 2).

Gene expression profiling of selected genes by q-RTPCR

Fourteen of 54 overexpressed genes identified after microarray analysis were selected for further gene expression validation: (1) *map1-6* ($\log_2\text{-FC} = 5.9$) and CDS_003900 (*hypo*) ($\log_2\text{-FC} = 8.2$) for their strong FC expression and (2) *ccmB*, *dksA*, *dapA*, *secF*, *sppA*, *atpB*, *cytoC1*, *trx*, *lolD*, *resolvase* (*resol*), *proP1*, and *folK*, mainly for their biological functions and involvement in the pathogenicity, the nutrient and protein transports, and the metabolism of the bacteria (Table 2). A typical profile of gene expression measured by q-RTPCR was shown for one experiment with Gardel p44 (Fig. 2c and d). The overexpression of *ccmB* ($\log_2\text{-FC} = 2.57$), *dksA* ($\log_2\text{-FC} = 4.07$), and *hypo*

($\log_2\text{-FC} = 5.06$) at EB stage and *map1-6* ($\log_2\text{-FC} = 1.88$), *dapA* ($\log_2\text{-FC} = 2.66$), *secF* ($\log_2\text{-FC} = 1.45$), and *sppA* ($\log_2\text{-FC} = 2.09$) at RB stage measured by q-RTPCR confirmed microarrays data (Fig. 2c and d, Table 2). *CcmB*, *dksA*, *hypo*, and *map1-6* were found overexpressed also in the two other experiments with Gardel p38 and p52. The strongest FCs were observed for *dksA* ($\log_2\text{-FC} = 8.79$ and 5.58) and *hypo* ($\log_2\text{-FC} = 6.45$ and 6.37) (data not shown).

For the other genes, an additional experiment with Gardel p41 was used either because there was a slight discrepancy in the gene expression trend or because there was no difference in expression between RB and EB stages for one of the three replicates. The overexpression of *dapA*, *secF*, and *sppA* genes was confirmed on this fourth experiment (data not shown). The expression of *resol*, *proP1*, and *folK* genes measured by q-RTPCR with four biological replicates did not confirm the overexpression observed by microarrays (data not shown). However, three independent replications of the time course showed that the remaining genes, *atpB*, *cytoC1*, *trx*, and *lolD*, were overexpressed at RB stage (data not shown). The mean $\log_2\text{-FC}$ obtained by q-RTPCR was 2.9 for *atpB*, 2.6 for *cytoC1*, 2.7 for *trx*, and 3 for *lolD* (data not shown).

Table 2. Overexpressed genes and functions determined by microarray analysis

Gene function	RB stage overexpression			EB stage overexpression		
	Gene id	Name	log ₂ -FC	Gene id	Name	log ₂ -FC
Energy production and conversion	CDS_05590	<i>nuoM</i>	2.6	CDS_05160	<i>cytoC1</i>[†]	4.6
	CDS_01340	<i>lpd</i>	6.1	CDS_08780	<i>atpB</i>[†]	4.6
Nucleotide transport and metabolism	CDS_07260	<i>dcd</i>	4.6	CDS_08290	<i>purK</i>	3.7
				CDS_05880		6.7
Coenzyme transport and metabolism	CDS_06730	<i>bioB</i>	1.44	CDS_03490	<i>coaD</i>	5.7
	CDS_06560	<i>dfp</i>	1.6	CDS_06750	<i>folK</i>[†]	6.2
	CDS_02710	<i>nadE</i>	4.3			
Replication, recombination, and DNA repair	CDS_07910		1.8	CDS_00420	<i>recF</i>	4.1
	CDS_08960	<i>recJ</i>	1.8	CDS_05820	<i>resol</i>[†]	4.7
Post-translational modification, protein turnover, chaperones	CDS_06350	<i>sppA</i>*	1.7	CDS_07850	<i>trx</i>[†]	3.2
				CDS_00340	<i>ccmb</i>*	4.9
				CDS_05090		5.5
Carbohydrate/AA/inorganic ion transport and metabolism	CDS_02670	<i>dapA</i>*	2.0	CDS_02740	<i>proP1</i>[†]	5.3
				CDS_07250		6.5
				CDS_04690	<i>tal</i>	5.0
Defense mechanisms				CDS_01120	<i>lold</i>[†]	5.5
Signal transduction mechanisms				CDS_00330	<i>dksA</i>*	8.2
Map1-related protein	CDS_09090	<i>map1-6</i>*	5.9			
Intracellular trafficking, secretion, and vesicular transport	CDS_00550	<i>secF</i>*	5.2			
Secondary metabolites' biosynthesis, transport, and catabolism	CDS_03920	<i>fabG</i>	2.0			
Translational, ribosomal structure, and biogenesis	CDS_05680	<i>rne</i>	3.3	CDS_05020	<i>rsme</i>	4.5
Hypothetical protein	CDS_00780		1.5	CDS_05770		3.0
	CDS_03690		2.2	CDS_01370		3.5
	CDS_04730		3.3	CDS_06790		3.6
	CDS_00080		5.1	CDS_09310		3.8
	CDS_08320		5.4	CDS_04750		3.9
				CDS_09210		4.2
				CDS_01730		4.4
				CDS_09340		4.5
				CDS_02510		5.2
				CDS_00640		5.8
				CDS_04560		5.9
				CDS_07620		6.0
				CDS_06420		6.8
				CDS_06710		6.9
				CDS_07360		7.0
				CDS_03620		7.3
				CDS_05100		7.6
				CDS_00390	<i>hypo</i>*	8.2

Bold text: selected genes for validation by q-RT-PCR; overexpression *Confirmed by q-RT-PCR. [†]Not confirmed by q-RT-PCR
FC: fold change.

Discussion

Up to now, few studies have been conducted on *Rickettsiales* transcriptomes (Leroy & Raoult, 2010). Moreover, functional studies of genes involved in the bacterial development and pathogenesis are challenging for obligate intracellular pathogens. Microarrays provide detailed knowledge of bacterial pathogenesis by high-throughput whole-genome analysis (Leroy & Raoult, 2010).

We assumed that the analysis of gene expression profiles from *ER* replicating inside bovine endothelial cells

would give clues on critical genes for *Ehrlichia* development and pathogenicity. We first used DNA microarrays to identify the genes differentially expressed during *ER* development. In a second time, we analyzed the expression profiles on selected genes by q-RT-PCR.

Obtaining *ER* synchronized and standardized cultures, as previously described by Marcelino *et al.* (2005) was crucial to ensure a good reproducibility between biological replicates. Despite our previous analysis on several genes described in the literature as internal reference genes for other pathogens (*16S*, *ffh*, *recA*, *rpoD*, and *proC*), we could

not identify constitutively expressed *ER* genes to use as normalizers for relative quantification of gene expression (data not shown). In this context, the gene expression was measured first by cDNA quantification for each targeted gene and then normalized by the number of bacteria per sample. Such method of normalization constitutes an optimal option for intracellular organisms (Vandecasteele *et al.*, 2002; Borges *et al.*, 2010).

Only 5% of CDS were identified as differentially expressed by microarrays between RB and EB stages. Hypothetical proteins represent the main functional category of overexpressed genes. It would be interesting to study the function of some of these *ER*-specific genes. We showed a differential expression of genes belonging to energy production and conversion, coenzyme transport and metabolism, replication, recombination, and DNA repair functional categories, with a higher proportion of genes at RB stage. *fabG* implicated in secondary metabolites' biosynthesis, transport, and catabolism and *rne* participating in translational, ribosomal structure and biogenesis, were also overexpressed at RB stage. These results are in accordance with the division phase of the bacteria, with a major transcriptional switch that correlated with the RB to EB transition (Leroy & Raoult, 2010). We found that genes involved in the carbohydrate, amino acid, inorganic ion, nucleotide, and coenzyme transports and metabolisms are differentially expressed at both RB and EB stages. However, *ER* was thought to have a condensed genome with a reduced gene expression at EB stage like *Chlamydia* (Nicholson *et al.*, 2003). On the contrary, our results suggest interestingly that EB of *Ehrlichia* could be metabolically active.

A differential expression was observed by q-RTPCR on 11 of 14 targeted genes, among which seven had different trend from microarrays data. This discrepancy could be due to the hybridization default of some microarrays probes. Similar expression profiles for the 11 genes were obtained between three biological replicates strengthening the accuracy of the q-RTPCR results, which were therefore considered to be the gold standard reference for further analysis.

After q-RTPCR analysis, the three genes, *atpB*, *cytoC1*, and *dapA* essential for the energy production and for the lysine biosynthesis, were found upregulated at RB stage (Domigan *et al.*, 2009). These data confirm the high metabolic activity, typical of RB of *ER* development.

At RB stage, we also showed an overexpression of *map1-6*, *secF*, *sppA*, and *lolD* genes that code for proteins involved in nutrient and protein exchanges and transports, indicating a possible important role in *ER* growth and division. *Map1-6* is essential for *ER* host adaptation and intracellular survival (Postigo *et al.*, 2008). The Sec system represents the major route in *Rickettsia typhi* for

protein secretion including the secretion of virulence factors (Ammerman *et al.*, 2008). *SppA* is the enzyme responsible for cleaving the signal peptide of Sec-dependent proteins. It is also described as a protease IV (Kim *et al.*, 2008; Golde *et al.*, 2009). Further study of this secretion system will give information of its involvement in *ER* virulence. *LolD* is a part of *LolCDE* protein complex that belongs to the ABC transporter superfamily and initiates the lipoprotein sorting to the outer membrane by catalyzing their release from the inner membrane. *LolCDE* complex is well conserved in various Gram-negative bacteria and thought to be essential for their growth (Narita & Tokuda, 2006).

The implementation of defense mechanisms against reactive oxygen species produced by host cells could be supported by the overexpression of *trx* we observed at RB stage. This gene contributes to the resistance to oxidative stress (Arner & Holmgren, 2000), and its overexpression has also been reported for *R. conorii* (Renesto *et al.*, 2008).

Three genes, *hypo*, *cmbB*, and *dksA*, were strongly overexpressed at EB stage. The *hypo* gene is identified in both *Ehrlichia canis* and *chaffeensis* and codes for an outer membrane protein, which is unique to the genus *Ehrlichia* (Miura & Rikihisa, 2007). The overexpression of this protein at EB stage suggests that it could be involved in the bacterium–host cell interaction and that a membrane reorganization could occur as described previously for *Chlamydia trachomatis* (Nicholson *et al.*, 2003). *CmbB* and *dksA* are localized on the same operon, thus suggesting that they could be coregulated. *CmbB* is known as a component of an ABC transporter involved in cytochrome C maturation (Richard-Fogal & Kranz, 2010). In *Salmonella typhimurium* and enterohemorrhagic *Escherichia coli*, *dksA* is involved in the virulence factor regulation, especially at late stage (Nakanishi *et al.*, 2006). In our model, EB may express virulence factors as *dksA* to improve new host cell invasion. Further investigations into the *dksA* function will be developed for *ER* model.

In conclusion, our study shows that several genes of *ER* are differentially expressed during the development stages. Here, we identified a number of known pathways as well as new genes that could be important for various aspects of *Ehrlichia* development and pathogenicity. Further analysis of these functions, in association with comparative genomic and proteomic approaches, will give us a better view of mechanisms of infection of *ER*.

Acknowledgements

The authors acknowledge the financial support received from European project, FEDER 2007–2013, FED 1/1.4–

30305, 'Risque en santé animale et végétale', and the Fundação para a Ciência e Tecnologia (FCT, Lisbon, Portugal; contract number PTDC/CVT/114118/2009). L.P. and L.E. acknowledge financial support for their PhD from the European project, FED 1/1.4-30305 and from 'Région Guadeloupe', respectively. I.M. acknowledges financial support from the grant SFRH/BPD/45978/2008 from FCT.

Authors' contribution

L.P., L.E., D.F.M. and N.V. contributed equally to this work.

References

- Allsopp BA (2010) Natural history of *Ehrlichia ruminantium*. *Vet Parasitol* **167**: 123–135.
- Ammerman NC, Rahman MS & Azad AF (2008) Characterization of Sec-translocon-dependent extracytoplasmic proteins of *Rickettsia typhi*. *J Bacteriol* **190**: 6234–6242.
- Arner ES & Holmgren A (2000) Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem* **267**: 6102–6109.
- Barre N, Uilenberg G, Morel PC & Camus E (1987) Danger of introducing heartwater onto the American mainland: potential role of indigenous and exotic *Amblyomma* ticks. *Onderstepoort J Vet Res* **54**: 405–417.
- Borges V, Ferreira R, Nunes A, Nogueira P, Borrego MJ & Gomes JP (2010) Normalization strategies for real-time expression data in *Chlamydia trachomatis*. *J Microbiol Methods* **82**: 256–264.
- Didier G, Brezellec P, Remy E & Henaut A (2002) GeneANOVA—gene expression analysis of variance. *Bioinformatics* **18**: 490–491.
- Domigan LJ, Scally SW, Fogg MJ *et al.* (2009) Characterisation of dihydroadipicinate synthase (DHAPS) from *Bacillus anthracis*. *Biochim Biophys Acta* **1794**: 1510–1516.
- Emboule L, Daigle F, Meyer DF *et al.* (2009) Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol Biol* **10**: 111.
- Frutos R, Viari A, Ferraz C *et al.* (2006) Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *J Bacteriol* **188**: 2533–2542.
- Golde TE, Wolfe MS & Greenbaum DC (2009) Signal peptide peptidases: a family of intramembrane-cleaving proteases that cleave type 2 transmembrane proteins. *Semin Cell Dev Biol* **20**: 225–230.
- Jongejan F, Zandbergen TA, van de Wiel PA, de Groot M & Uilenberg G (1991) The tick-borne *rickettsia* *Cowdria ruminantium* has a *Chlamydia*-like developmental cycle. *Onderstepoort J Vet Res* **58**: 227–237.
- Kim AC, Oliver DC & Paetzel M (2008) Crystal structure of a bacterial signal Peptide peptidase. *J Mol Biol* **376**: 352–366.
- Leroy Q & Raoult D (2010) Review of microarray studies for host-intracellular pathogen interactions. *J Microbiol Methods* **81**: 81–95.
- Marcelino I, Verissimo C, Sousa MF, Carrondo MJ & Alves PM (2005) Characterization of *Ehrlichia ruminantium* replication and release kinetics in endothelial cell cultures. *Vet Microbiol* **110**: 87–96.
- Martinez D, Vachieri N, Stachurski F, Kandassamy Y, Raliniaina M, Aprelon R & Gueye A (2004) Nested PCR for detection and genotyping of *Ehrlichia ruminantium*: use in genetic diversity analysis. *Ann N Y Acad Sci* **1026**: 106–113.
- Miura K & Rikihisa Y (2007) Virulence potential of *Ehrlichia chaffeensis* strains of distinct genome sequences. *Infect Immun* **75**: 3604–3613.
- Nakanishi N, Abe H, Ogura Y *et al.* (2006) ppGpp with DksA controls gene expression in the locus of enterocyte effacement (LEE) pathogenicity island of enterohaemorrhagic *Escherichia coli* through activation of two virulence regulatory genes. *Mol Microbiol* **61**: 194–205.
- Narita S & Tokuda H (2006) An ABC transporter mediating the membrane detachment of bacterial lipoproteins depending on their sorting signals. *FEBS Lett* **580**: 1164–1170.
- Nicholson TL, Olinger L, Chong K, Schoolnik G & Stephens RS (2003) Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. *J Bacteriol* **185**: 3179–3189.
- Postigo M, Taoufik A, Bell-Sakyi L, Bekker CP, de Vries E, Morrison WI & Jongejan F (2008) Host cell-specific protein expression *in vitro* in *Ehrlichia ruminantium*. *Vet Microbiol* **128**: 136–147.
- Renesto P, Rovey C, Schrenzel J *et al.* (2008) *Rickettsia conorii* transcriptional response within inoculation eschar. *PLoS ONE* **3**: e3681.
- Richard-Fogal C & Kranz RG (2010) The CcmC:heme:CcmE complex in heme trafficking and cytochrome c biosynthesis. *J Mol Biol* **401**: 350–362.
- Vandecasteele SJ, Peetermans WE, Merckx R, Van Ranst M & Van Eldere J (2002) Use of gDNA as internal standard for gene expression in staphylococci *in vitro* and *in vivo*. *Biochem Biophys Res Commun* **291**: 528–534.

2.4. Analyse du transcriptome des souches Gardel et Sénégal virulente et atténuée au stade corps élémentaire (T3). Transcriptome of virulent and attenuated strains of *Ehrlichia ruminantium* reveal important genes in attenuation mechanisms (article en préparation).

A partir des nouvelles données de microarrays, une première analyse a été effectuée sur les échantillons à T3 sans captures pour l'ensemble des souches. Les résultats de coefficients de corrélation linéaire obtenus montraient une bonne homogénéité entre ces différents réplicats. Au sein d'une même souche (Gardel Sénégal virulente ou atténuée), nous avons moyenné pour chaque sonde les intensités de fluorescence obtenues pour les différents réplicats.

Nous avons identifié d'une part les gènes différentiellement exprimés entre les souches Gardel virulente et atténuée et d'autre part, les gènes différentiellement exprimés entre les souches Sénégal virulente et atténuée. Dans un second temps, les listes obtenues des gènes ont été ensuite confrontés afin d'identifier les gènes communs à Gardel et à Sénégal.

Pour les souches Gardel, il y avait 58 et 64 gènes sur-exprimés chez la souche virulente et chez la souche atténuée respectivement. Ces gènes codaient majoritairement pour des protéines hypothétiques. Chez les souches Sénégal, le nombre de gènes sur-exprimés était plus important chez la virulente. Ce nombre était de 76 contre 51 chez l'atténuée et comme pour Gardel avec une forte proportion de gènes hypothétiques.

L'ensemble des gènes obtenu a été assigné par groupes fonctionnels, cela nous a permis ensuite d'identifier des groupes fonctionnels de gènes spécifiques d'une ou de plusieurs souches. Par exemple, les gènes codant pour les protéines MAP et les gènes impliqués dans les mécanismes de transduction de signaux, étaient trouvés sur-exprimés chez les deux souches virulentes Gardel et Sénégal. Il n'y avait pas de groupes fonctionnels spécifiques des souches atténuées, mais la proportion de gènes impliqués dans les fonctions de métabolisme de base était globalement plus

importante que chez les souches virulentes. Par exemple, les gènes impliqués dans la production d'énergie représentaient 2 et 3% des gènes sur-exprimés chez Gardel et Sénégal virulentes respectivement tandis qu'ils représentaient 5% chez Gardel atténuée et 12% chez Sénégal atténuée. Nous avons également observé une sur-expression des gènes codant pour le SST4. En effet, 10 gènes essentiellement localisés sur les deux opérons majeurs du SST4 étaient trouvés sur-exprimés chez la souche Gardel virulente. Comme pour les gènes de la famille multigénique *map1*, il semble y avoir une co-régulation des gènes de ce système de sécrétion.

Enfin, nous avons également montré une expression différentielle de plusieurs gènes impliqués dans la lutte contre les mécanismes de défense de la cellule hôte (stress oxydatif et osmotique). Les gènes impliqués dans la lutte contre le stress oxydatif étaient trouvés uniquement chez les souches Sénégal. Ils étaient au nombre de 6 et répartis équitablement entre les deux souches. Ce résultat semble indiquer une stratégie de lutte différente entre la souche virulente et la souche atténuée.

Dans un second temps, pour l'ensemble des gènes différentiellement exprimés, nous avons vérifié la présence de mutations éventuelles. 16 gènes étaient trouvés tronqués ou divisés entre les différentes souches. La mutation de l'un d'entre eux *recO*, a été montrée comme impliquée dans le mécanisme d'atténuation de la *Rickettsiales R. prowazekii*.

Transcriptome of virulent and attenuated strains of *Ehrlichia ruminantium* reveal important genes in attenuation mechanisms

Ludovic Pruneau^{1,2}, Bernard Mari³, Isabel Marcelino^{1,4}, Kevin Lebrigand³, Alain Viari⁵, Loïc Emboulé¹, Valérie Pinarello¹, Christian Sheikboudou¹, Thierry Lefrançois⁵, Damien F Meyer¹ and Nathalie Vachiery^{1*}

¹ CIRAD, UMR CMAEE, F- 97170 Petit-Bourg, Guadeloupe, France

² Université des Antilles et de la Guyane, Fouillole, BP-250 97157 Pointe-à-Pitre, Guadeloupe, France

³ UMR 6097, CNRS-Université de Nice Sophia Antipolis, Institut de Pharmacologie Moléculaire et Cellulaire, Sophia Antipolis, F06560, France

⁴ IBET, Apartado 12, 2780-901 Oeiras, Portugal

⁵ CIRAD, UMR CMAEE, F-34398 Montpellier, France

*corresponding author: e-mail: nathalie.vachiery@cirad.fr

Tel: +590-590255442, Fax: +590-590940396

Keywords: *Ehrlichia*, transcriptomic, pathogenicity,

Abstract

Ehrlichia ruminantium is the causal agent of Heartwater, a fatal disease in ruminants. This obligate intracellular bacterium is transmitted by ticks of the genus *Amblyomma*. Here, we report the global transcriptome of 4 strains of *E. ruminantium*, Gardel and Senegal, two distant virulent strains of *E. ruminantium* with their corresponding attenuated strains. Our results showed a higher metabolic activity in attenuated strains compared to virulent strains, suggesting a better adaptation *in vitro* of attenuated strains to the host cells. There was a strong modification of membrane protein expression for attenuated strains. An exclusive over-expression of *map1*-related genes, was observed for virulent strains, whereas attenuated strains over-expressed genes encoding for hypothetical membrane proteins. This result suggests that *in vivo*, MAP-1 related proteins could induce non-protective immune responses for virulent strains. For the attenuated strains, the lack of expression of *map1*-related genes and over-expression of other membrane proteins could be important in induction of efficient immune responses.

The diminution of expression of many genes in attenuated Senegal was caused by severe mutation. One of them, the gene *recO* is involved in DNA repair and its mutation could explain the higher proportion of mutated genes in attenuated Senegal, unducing the faster attenuation of Senegal compared to Gardel.

Introduction

The *Rickettsiales* order includes small obligate intracellular α -proteobacteria, most closely related to mitochondria ([Merhej & Raoult, 2011](#)). These bacteria are transmitted by blood-sucking arthropods, and are responsible for major endemic and emerging human and animal infectious diseases with important economical and sanitary impacts.

Due to their obligatory intracellular characteristic, the understanding of the *Rickettsiales* biology was strongly hampered, resulting in culture constraints and difficulties to study the genetics of these bacteria. Moreover, *Rickettsiales* mechanisms of virulence and attenuation are not yet completely elucidated. *Rickettsiales* members establish mechanisms for survival and development in the host cells, such as the escape to osmotic and oxidative stress. Moreover, outer membrane proteins play a key role in interaction of *Rickettsiales* with their host/vector cells.

E. ruminantium is the causal agent of heartwater, a fatal disease in wild and domestic ruminants ([Allsopp, 2010](#)). This disease represents a serious problem for livestock productivity in endemic areas such as sub-Saharan Africa and the Caribbean threatening the American continent where indigenous competent ticks are present ([Barre, et al., 1987](#)). *E. ruminantium* belongs to the order *Rickettsiales* and the *Anaplasmataceae* family, which contains others genera, such as *Anaplasma*, and *Neorickettsia*. *E. ruminantium* is transmitted by *Amblyomma* ticks, mainly *A. variegatum* present in Africa and Caribbean islands and *A. hebraeum* present in the Southern of Africa. Improvement of *in vitro* culture conditions and availability of new *E. ruminantium* genome sequences allow using high throughput integrative approaches like genomic, transcriptomic and proteomic to understand the pathogenesis of *E. ruminantium* and most widely of *Rickettsiales*.

We report the differential gene expression between virulent and attenuated Gardel and Senegal strains. Gardel is a strain isolated in Guadeloupe (French West Indies) and Senegal is strain originating from Senegal. These strains are phylogenetically distant ([Raliniaina, et al., 2010](#), [Pilet, et al., 2012](#))

with different behavior *in vitro* and different attenuation speed. Indeed, *in vitro*, Gardel is attenuated after 200 passages whereas Senegal is attenuated only after 15 passages. The transcriptomic analyze was performed on elementary body of strains, that is infectious form.

In this study, we propose to better understand the mechanisms of virulence and attenuation of *E. ruminantium* by comparison of global transcriptome between virulent and attenuated strains.

Materials and Methods

Production of biological replicates

Three biological replicates for each strain were used for the microarrays experiments. For virulent and attenuated Gardel, biological samples were produced at passage 38, 43 and 44 and passage 238, 243 and 251 respectively. For virulent Senegal, the replicates were Senegal passage 7, and two replicates passage 8. For attenuated Senegal, biological samples were Senegal passages 70, 75 and 77. All the biological samples were produced in bovine aorta endothelial cells (BAE) as described by ([Emboule, et al., 2009](#)).

When 80% cell lysis was observed, supernatant and cellular debris were harvested and then ultra-centrifuged at $20,000 \times g$ for 15 min at 4°C to collect elementary bodies (EB). The pellets were placed in sterile eppendorfs and homogenized in 2.5 ml of TRIzol reagent (Invitrogen). The cells lysed immediately and released RNA and DNA in the supernatant. The samples were immediately stored at -80°C before RNA extraction.

Extraction of total RNA

Total RNA extraction procedure was carried out as described in ([Emboule, et al., 2009](#)). Total RNA quantification was performed by Nanodrop 2000c

(Thermo Scientific), and total RNA samples were pooled in RNase free water at a final concentration of 0.5 µg µL⁻¹.

Reverse transcription of RNA samples

Before the reverse transcription of RNA samples, the bacterial gDNA contaminant was evaluated by PCR 25 cycles targeting *pCS20* gene using primers AB128 and AB129 as described previously ([Martinez, et al., 2004](#)). In RNA samples, no signal was obtained by *pCS20* PCR (data not shown). Afterward, RNA samples were reverse transcribed by random priming with Superscript II (Invitrogen) according to manufacturer's instructions. The reverse transcription and PCR amplification of corresponding cDNA were done using *KpnI*-primers as previously described ([Emboule, et al., 2009](#)). Moreover, the efficiency of conversion was checked using q-RTPCR targeting *ER* 16S gene by processing simultaneously RNA and cDNA samples ([Emboule, et al., 2009](#)). The difference of Ct between RNA and cDNA samples was always higher than 5 cycles indicating then low contamination by gDNA in RNA samples (data not shown).

Design of *ER* microarrays

ER microarray (8 x 60 k) used in this study were manufactured by Agilent, France. The probes were designed with a modified version of Oligoarray program (Rouillard, et al., 2003). Design of probes was carried out from genomes of 3 *ER* strains published (virulent Gardel, Welgevonden Erwe and Erwo) ([Collins, et al., 2005](#), [Frutos, et al., 2006](#)) and from genomes of 3 *E. ruminantium* strains newly sequenced but not published (attenuated Gardel, virulent and attenuated Senegal). There were common probes to 6 strains and probes specific to one or more strains. Probes designed had a size of 60-mer, a Tm between 70 and 92°C and a GC percent between 20 and 60%.

The microarrays contained a total of 16 199 probes, with 3 replicates per probe, including 83 bovine genes as negative controls. Experimental data and associated microarray designs have been deposited in the NCBI Gene

Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/geo/> under platforms.

Microarrays expression and analysis

Two hundreds nanograms of cDNA were randomly amplified and labeled with Cy3-dUTP using the SureTag DNA Complete Labeling Kit (Invitrogen, France). Following purification with reaction purification columns (Invitrogen, France), quantification of Cy3-dUTP incorporation was performed by absorbance measurement at 550 nm. The yield of cDNA labeling ranged between 3 and 5 µg, and the specific activity ranged between 20 and 25 pmol per µg of cDNA, according to the manufacturer's instructions (Invitrogen). Microarrays were incubated for 24 h at 65°C, with a rotation of 20 rpm, in a hybridization chamber. After hybridization, stringent washings were performed according to manufacturer's instructions (Invitrogen). Arrays were scanned using Agilent C microarray Scanner and extracted with Agilent Feature Extraction program (version 10.10). Images obtained, were saved in TIFF format and data with the signal intensities of all spots on each image were saved as ".txt" files for further analysis. Microarray results were analyzed with GeneANOVA software ([Didier, et al., 2002](#))

Before the statistical analysis, homogeneity of log2 transformed fluorescence intensities between biological replicates was checked by computing linear correlation coefficients. For virulent and attenuated Gardel and Senegal strains, linear correlation coefficients must be greater than 0.6. The mean intensity values of biological replicates were then calculated.

Identification of differential gene expressed

Fold change (FC) between virulent and attenuated strains were calculated for each probe. Genes were considered differentially expressed between virulent and attenuated strains for Gardel and Senegal when more than 60%

of probes have a \log_2 -fold change (FC) > 0.9 and a p-value < 0.1. A mean of FC per CDS was calculated.

The annotations of genes differentially expressed were checked on NCBI. For hypothetical genes, proteins patterns were searched in the database InterProt (Quevillon *et al.*, 2005) and alignments were performed by BLAST (Basic Local Alignment Search Tool) to identify potential functional domains. For each differentially expressed gene, its sequence in the attenuated strain was aligned with the sequence in the virulent one in order to identify potential mutations or deletions.

Results

Reproducibility of the transcriptome profiles for biological replicates

The biological replicates of Gardel and Senegal virulent and attenuated strains were hybridized on microarrays. The signal intensities obtained were compared between the different biological replicates and the linear correlation coefficients between replicates were calculated for each strain. As illustrated by figure 1-A, linear correlation coefficients between the different biological replicates were higher than 0.6. Furthermore, whole-genome-wide transcript expression profile analysis for different replicates were compared between them and a clustering was then carried out with the 300 best probes discriminating virulent and attenuated strains. One hundred and ninety six probes were differentially expressed between virulent and attenuated Gardel and 91 probes for Senegal. Only one probe was common to Gardel and Senegal. Clusters based on probe intensity allowed identifying 4 groups each group including biological triplicates. Virulent Gardel group is clearly separated from other groups formed by attenuated Gardel and attenuated and virulent Senegal (figure 1-B).

General overview of overexpressed gene functions for Gardel and Senegal strains

There were 122 and 127 out of 950 genes (12.84% and 13.37% of ORFs) differentially expressed between virulent and attenuated Gardel and Senegal strains respectively. Fifty-eight genes were found overexpressed for virulent Gardel strain, and 64 genes for attenuated Gardel strain with a log₂-FC ranging from 0.94 to 3.84 (Table 1). Compare to Gardel, there were more genes up regulated in Senegal virulent strain (76 genes) than attenuated strain (51 genes), with lower log₂-FC ranging from 0.97 to 2.93 (Table 2).

These genes were classified according to their COG (clusters of orthologous groups of proteins) function. The proportion of overexpressed gene categories for virulent and attenuated Gardel strains was shown in figure 2A and 2B, and for virulent and Senegal strains, in figure 3A and 3B respectively. Genes with hypothetical functions represented the highest proportion of overexpressed genes both for Gardel and Senegal. There were 52% of overexpressed hypothetical genes for virulent Gardel strains including 28% of hypothetical genes, 22% of hypothetical membrane genes and 2% of hypothetical exported genes. For attenuated Gardel strain, the percentage of hypothetical genes was 45% with 23% of hypothetical genes, 20% of hypothetical membrane genes and 2% of hypothetical exported genes. For Senegal, there were 44% and 41% of hypothetical genes overexpressed for virulent and attenuated strains respectively (Figure 3A and 3B), but with a notable difference for exported hypothetical protein. Indeed, the proportion of genes encoding for exported hypothetical protein were higher for Senegal, there were 8% for virulent Senegal and 4% for attenuated Senegal compared to 2% for Gardel.

More interestingly, genes encoding for Map1-related protein and for proteins involved in signal transduction mechanisms were overexpressed exclusively in virulent strains. Map1-gene family represented 14% and 12% of overexpressed genes for Gardel and Senegal virulent strains respectively. There was no over-expression of genes involved in signal

transduction mechanisms for attenuated strains and 3% of overexpressed genes for virulent Gardel and Senegal strains. Intracellular trafficking, secretion and vesicular transport were exclusively overexpressed for virulent Gardel (17%) due to 10 genes encoding for components of type 4 secretion system (T4SS) and for attenuated Senegal (6%) due to 3 genes, *virB3*, *virB6a* and *lepB* (Table 2). Genes involved in transcription were also up-regulated in virulent Gardel and attenuated Senegal (figure 2A and 3B). There were just 2 COG specifically up-regulated for one strain. COG of defense mechanisms was specific for attenuated Gardel (figure 2A) and COG of secondary metabolites biosynthesis, transport and catabolism was specific for attenuated Senegal (figure 3A).

Furthermore, genes involved in nucleotide transport and metabolism, cell wall/membrane/envelope biogenesis, and translation, ribosomal structure and biogenesis were overexpressed for attenuated Gardel and for both virulent and attenuated Senegal but not for virulent Gardel. Interestingly, COG corresponding to Replication, recombination and repair was found for all strains except for attenuated Senegal.

Finally, COGs of genes corresponding to basic metabolic functions such as Energy production and conversion, Carbohydrate/ Amino acid/ Inorganic ion transport and metabolism, Coenzyme transport and metabolism, and to Posttranslational modification, protein turnover, chaperones, were common to the four strains, but globally the proportions of genes overexpressed in these COGs were higher in attenuated strains (figure 2A, 2B, 3A and 3B).

Genes encoding membrane proteins are over-expressed

Over-expression of *map1* gene family exclusively in virulent strains

Among the 16 genes encoding for Map1-related protein, 11 genes were up-regulated for virulent Gardel and Senegal strains compared to attenuated strains (Figure 4). The over-expression was common for virulent Gardel and Senegal for *map1-12*, *map1-9*, *map1-8*, *map1-7*, *map1-6* and *map1-4*. *Map1-13*, *map1-10*, and *map1-2* were only overexpressed in virulent Senegal strain

whereas *map1-5* and *map1-3* were only overexpressed in Gardel strain. FC-log₂ of *map1-9* (2.55 vs. 1.82), *map1-8* (1.56 vs. 1.39) and *map1-7* (2.08 vs. 1.50) were higher for Senegal compared to Gardel.

Over-expression of genes encoding for hypothetical membrane proteins

Genes encoding for hypothetical membrane proteins were found for the 4 strains. The proportions of overexpressed genes were similar for virulent and attenuated Gardel (22% vs 20%). For Senegal strains, the proportions were lower compared to Gardel, 16% and 10% for virulent and attenuated Senegal respectively. Moreover, there were 4 genes in common for the 4 strains. CDS_03690 was commonly overexpressed in attenuated Gardel and Senegal strains whereas both CDS_02900 and CDS_03260 were overexpressed in virulent Senegal and attenuated Gardel, and CDS_07940 was overexpressed in virulent Gardel and attenuated Senegal. Except CDS_03260, these genes are specific of *E. ruminantium*.

Differential expression of T4SS components between virulent and attenuated Gardel strains

In *E. ruminantium*, genes encoding for T4SS are organized in 5 clusters comprising 2 major operons. Ten *vir* genes were overexpressed exclusively in Gardel virulent with a higher log₂-FC (3.72 to 1.39, Table 1). There were *virB4a*, *virB6a*, *virB6b*, *virB6c*, and *virB6d* localized on major operon *virB3-B6*; *virB10*, *virB11*, and *virD4* localized on second major operon *virB8-11/D4*; and *virB2b* and *virB2c* localized on operon *virB4b/B2a-d*.

For Senegal, there were only 2 *vir* genes (*virB3* and *virB6a*) differentially expressed between virulent and attenuated strains, but contrary to Gardel, there were overexpressed for attenuated strain.

***E. ruminantium* response against defense mechanisms of host cells**

Among the genes differentially expressed between virulent and attenuated Gardel and Senegal strains, several are involved in host cell response

escape. In fact, *grxC2* which encodes for glutaredoxin was overexpressed both in attenuated Gardel and Senegal strains. *Trx* and *bcp* encoding for thioredoxin and for Bacterioferritin comigratory protein respectively, were also overexpressed but only in attenuated Gardel strains. Only Senegal displayed differential gene expression for genes involved in counteracting osmotic stress. For Senegal, there were 6 genes differentially expressed between virulent and attenuated strains. *ProP* (proline-betaine transporter), *nuoE* (NADH dehydrogenase subunit E) and CDS_01720 (Putative Na⁺/H⁺ antiporter subunit) were overexpressed for virulent strain. *NuoH* (NADH dehydrogenase subunit H), *nuoK* (NADH-quinone oxidoreductase subunit K) and CDS_05720 (Putative Na⁺/H⁺ antiporter subunit) were overexpressed for attenuated strain.

Others genes known to be induced during stress, were found to be differentially expressed between virulent and attenuated strains with a high log₂-FC. For example, genes encoding for ClpB (2.59) protein and for the heat shock protein GrpE (1.65) were overexpressed in attenuated Senegal. Moreover, two others genes encoding for BolA-like protein were also overexpressed for attenuated Senegal (CDS_07190) and Gardel (CDS_01800) strains.

Mutation in genes with altered expression

Comparison of genomic sequences of genes differentially expressed between virulent and attenuated strains revealed severe mutations inducing gene truncature or splitting. Two hypothetical genes overexpressed in virulent Gardel, CDS_04510 and CDS_08730, are split in attenuated Gardel. CDS_08730 encodes for a Pentapeptide repeats (TPR)-containing protein. For Senegal strain, 8 genes differentially expressed were split or truncated in the attenuated strain. One of these gene (*recO*) encodes for a component of complex RecFOR involved in DNA repair. *RpoH* Gene encoding for RNA polymerase sigma-32 factor, was overexpressed in attenuated Senegal, but truncated for this strain and split in virulent Senegal strain. Another gene

CDS_07940 that encodes for a hypothetical membrane protein is truncated in Senegal strains and attenuated Gardel strain. This gene was over-expressed for attenuated Gardel and virulent Senegal strains.

Discussion

Identification of pathogenic mechanisms is crucial for the understanding of biology of pathogenic bacteria. In *Rickettsiales* order, the majority of studies have been conducted only on virulent strains. In this study, we compared the global transcriptome of two distant virulent strains of *E. ruminantium* with their corresponding attenuated strains. Moreover, three biological replicates by strain were used thus strengthening our microarrays data.

Replicates of Gardel virulent were distant from other (figure 1-B). This result is correlated with differences of COGs between the 4 strains. In fact, 3 COGs were absent in virulent Gardel compared to other strains. One of these COGs contains genes involved in translation, ribosomal structure and biogenesis. For Gardel virulent, the absence of this COG is surprising but suggests that elementary bodies of Gardel virulent have a lower protein synthesis and enzymatic activity. Our recent study on the transcriptome of reticulate and elementary bodies of Gardel virulent, indicate a higher enzymatic activity more important in reticulate body compared to elementary body ([Pruneau, et al., 2012](#)). It will be interesting to compare the gene expression involved in the enzymatic activity and metabolism at the other stages of development between virulent and attenuated strains.

Globally, attenuated strains seem to have a higher metabolic activity than virulent strains. This result suggests a better adaptation of attenuated strains for the *in vitro* culture and could be linked to the shorter development cycle of attenuated strains.

Like many intracellular pathogen, *E. ruminantium* seems to be able to fight efficiently against osmotic and oxidative stresses generated by host cell. Osmotic and oxidative stresses are key methods by which mammalian

infected cells kill bacteria. Genes involved in counteracting oxidative stress were found exclusively for attenuated strains which could explain their better development *in vitro*. Gene encoding for glutaredoxin (*grxC2*) was overexpressed for the two attenuated strains. This over-expression of *grxC2* was reported also for *Rickettsia conorii* infected eschars ([Renesto, et al., 2008](#)). Another gene involved in fight against oxidative stress *trx*, was found to be over-expressed in elementary body of virulent Gardel compared to reticulate body ([Pruneau, et al., 2012](#)). *Trx* was found over-expressed for elementary body of attenuated Gardel strain. The modulation of gene expression along the infection for virulent and attenuated strains will be studied in further analysis. It seems that attenuated strains have a higher activation of mechanisms against oxidative stress. Only Senegal strains had differential expression of genes involved in fight against osmotic stress. Three genes were over-expressed for each strain indicating a different activation of these genes between attenuated and virulent Senegal. For Gardel strains, there were no genes differentially expressed and involved in fight against osmotic stress. This result suggests that virulent and attenuated Gardel strains could use common mechanisms to fight against osmotic stress with the activation of similar genes.

Genes involved in signal transduction mechanisms were over-expressed exclusively in virulent strains. One gene involved in signal transduction mechanisms *dksA* encodes for a transcriptional factor and found to be over-expressed for virulent Gardel. This gene was also over-expressed for elementary body compared to reticulate body ([Pruneau, et al., 2012](#)). In *Salmonella typhimurium* and enterohemorrhagic *Escherichia coli*, *dksA* regulates virulence factors especially at late stage ([Nakanishi, et al., 2006](#)). It seems that at late stage of development, *E. ruminantium* activates virulence genes for a new infectious cycle.

Genes encoding for Map1-related proteins were also exclusively over-expressed in virulent strains. These proteins are major antigenic proteins and in the case of MAP1 induce a strong humoral response, which is not

protective. Previous studies shown that *map1* family genes were differentially expressed in host or vector cells suggesting that they are involved in the adaptation of *E. ruminantium* to its host or vector ([Postigo, et al., 2008](#)). The exclusive over-expression of genes encoding for MAP1-related proteins in virulent strains suggests that they could induce a non-protective immune responses. For attenuated strains, many genes encoding for hypothetical membrane protein, as well as genes involved in posttranslational modifications and cell wall/membrane/envelope biogenesis were over-expressed. This result suggests membrane reorganization as described previously for *Chlamydia trachomatis* ([Nicholson, et al., 2003](#)). The lack of expression of *map1* family genes in attenuated strains and the over-expression of other membrane proteins could be important in induction of efficient immune responses.

Genes from two major T4SS clusters were found over-expressed in virulent Gardel compared to attenuated Gardel. T4SS is important for pathogenesis of intracellular bacteria because it permits translocation of virulence factors into the host cell. Over-expression of genes encoding for T4SS suggests an important role in the virulence of *E. ruminantium*.

Among genes differentially expressed between virulent and attenuated strains, certain are truncated or split in attenuated strains. One of these genes, *recO* which is over-expressed in virulent Senegal strain, is truncated in attenuated Senegal strain. This gene encodes for a protein of RecFOR complex that is important for DNA reparation. Moreover, in a recent study in another *Rickettsiales*, *Rickettsia prowazekii*, the authors showed that inactivation of *recO* resulted in a loss of virulence of *R. prowazekii* ([Bechah, et al., 2010](#)). Mutation of *recO* in *E. ruminantium* could explain the higher proportion of mutated genes in attenuated Senegal as compared to attenuated Gardel. This could explain the faster attenuation of Senegal compared to Gardel.

In conclusion, the comparison of global transcriptomes of virulent and attenuated strains suggested a reorganization of the membrane of *E.*

ruminantium. The overexpressed membrane proteins expressed in attenuated strains could be vaccine candidates against heartwater and their function will be studied in detail. We also showed the potential impact of mutations in the attenuation mechanisms. Future investigations will be conducted on the proteome of *E. ruminantium* in order to confirm our transcriptomics data. The genome comparison of the 4 strains should help us to identify other mutated genes involved in mechanisms of attenuation.

Table 1: Overexpressed genes and functions for Gardel strains

Gene function	virulent			attenuated		
	Gene id	Name	log ₂ FC	Gene id	Name	log ₂ FC
Energy production and conversion	CDS_05160	<i>Cytochrome c1</i>	2,6	CDS_04300	<i>fdxB</i>	1,99
				CDS_00110	<i>fdxA</i>	1,58
				CDS_08790	<i>atpE</i>	1,16
Nucleotide transport and metabolism				CDS_02940	<i>dnaX</i>	1,86
				CDS_07770	<i>tmk</i>	1,62
				CDS_08980	<i>ndk</i>	1,25
				CDS_06990	<i>gmk</i>	1,23
Coenzyme transport and metabolism	CDS_00530	<i>hemA</i>	1,20	CDS_00720	<i>ribB</i>	1,59
				CDS_07700	<i>ribE</i>	1,55
				CDS_01790	<i>pdxH</i>	1,46
Replication, recombination and DNA repair	CDS_03430	<i>topA</i>	2,19	CDS_05820	<i>resolvease</i>	1,88
	CDS_06660	<i>radC</i>	1,46	CDS_02430	<i>nth</i>	1,39
				CDS_02840	<i>ssb</i>	1,27
Posttranslational modification, protein turnover, chaperones	CDS_04780	<i>clpA</i>	1,79	CDS_05010	<i>bcp</i>	1,83
	CDS_07180	<i>dsbE</i>	1,61	CDS_07200	<i>grxC2</i>	1,34
				CDS_07850	<i>trx</i>	1,26
Carbohydrate/AA/inorganic ion transport and metabolism	CDS_03250	<i>N-Glu deshydro</i>	1,73	CDS_05860	<i>thio</i>	1,92
				CDS_04200	<i>rpiB</i>	1,6

Defense mechanisms	CDS_01120	<i>lolD</i>	1,52
Lipid transport and metabolism	CDS_05660		1,28
Signal transduction mechanisms	CDS_03230		1,25
	CDS_00330	<i>dksA</i>	1,16
Transcription	CDS_01740	<i>coaX</i>	1,37
Cell wall/ membrane/ envelope biogenesis	CDS_08860	<i>lgt</i>	1,89
	CDS_08610	<i>dacF</i>	1,62
	CDS_06490	<i>aprD</i>	1,49
	CDS_03590	<i>lepA</i>	1,39
	CDS_01510	<i>aprE</i>	1,31
Intracellular trafficking, secretion, and vesicular transport	CDS_08380	<i>virB2b</i>	3,72
	CDS_08370	<i>virB2c</i>	2,62
	CDS_00150	<i>virB10</i>	2,19
	CDS_00140	<i>virB11</i>	1,74
	CDS_00130	<i>virD4</i>	1,39
	CDS_05400	<i>virB4a</i>	1,51
	CDS_05390	<i>virB6a</i>	1,42
	CDS_05380	<i>virB6b</i>	2,11
	CDS_05370	<i>virB6c</i>	1,92
	CDS_05360	<i>virB6d</i>	1,67
Map1-related protein	CDS_09120	<i>map1-3</i>	2,21
	CDS_09110	<i>map1-4</i>	2,06
	CDS_09090	<i>map1-6</i>	1,84
	CDS_09060	<i>map1-9</i>	1,82
	CDS_09030	<i>map1-12</i>	1,59
	CDS_09100	<i>map1-5</i>	1,53
	CDS_09080	<i>map1-7</i>	1,50
	CDS_09070	<i>map1-8</i>	1,39
Translation, ribosomal structure and biogenesis	CDS_05070	<i>rrmJ</i>	2,50
	CDS_05020	<i>RsmE</i>	2,18
	CDS_00280	<i>trmE</i>	1,70
	CDS_06320		1,53
	CDS_05240	<i>rpsB</i>	1,22
	CDS_06080	<i>rpsM</i>	1,20

			CDS_03560	<i>rpsO</i>	1,10
			CDS_09290	<i>rplS</i>	1,14
			CDS_06190	<i>rplN</i>	1,09
			CDS_04800	<i>rbfA</i>	1,13
Hypothetical membrane protein	CDS_07450	2,81	CDS_06520		2,57
	CDS_07370	2,65	CDS_02270		2,32
	CDS_04750	2,56	CDS_00970		2,30
Hypothetical membrane protein	CDS_07440	2,41	CDS_02290		2,22
	CDS_07630	2,36	CDS_07620		2,22
	CDS_07670	2,36	CDS_03690		1,98
	CDS_07610	2,29	CDS_04760		1,85
	CDS_02390	2,26	CDS_05900		1,60
	CDS_08500	2,15	CDS_08300		1,53
	CDS_06460	2,04	CDS_07570		1,50
	CDS_05650	1,73	CDS_07940*		1,46
	CDS_03260	1,37	CDS_07470		1,32
	CDS_02900	1,20	CDS_03610		1,24
Hypothetical exported protein	CDS_08320	2,16	CDS_05140		3,84
Hypothetical protein	CDS_01030	3,81	CDS_07600		3,74
	CDS_05600	3,69	CDS_09210		2,46
	CDS_04510*	3,68	CDS_05350		2,24
	CDS_05770	2,76	CDS_06720		1,98
	CDS_08310	2,73	CDS_04190		1,81
	CDS_02370	2,59	CDS_03280		1,76
	CDS_05620	2,55	CDS_08200		1,76
	CDS_02410	2,37	CDS_01260		1,59
	CDS_08100	1,94	CDS_05640		1,58
	CDS_00600	1,87	CDS_03450		1,44
	CDS_02500	1,79	CDS_01780		1,20
	CDS_00570	1,79	CDS_04680		1,19
	CDS_08730*	1,68	CDS_02440		1,10
	CDS_02830	1,50	CDS_09300		0,97
	CDS_07680	1,40	CDS_01800		0,94
	CDS_02340	1,36			

*gene with mutation on attenuated strain, **Bold:** genes identified as differentially expressed in both Gardel and Senegal strains.

Table 2: Overexpressed genes and functions for virulent Senegal strains

Gene function	virulent			attenuated		
	Gene id	Name	log ₂ FC	Gene id	Name	log ₂ FC
Energy production and conversion	CDS_04540	<i>nuoE</i>	1,59	CDS_04410	<i>nuoH</i>	1,87
	CDS_01840	<i>sdhC</i>	1,55	CDS_04070	<i>atpG</i>	1,83
				CDS_04300	<i>fdxB</i>	1,76
				CDS_04700	<i>atpC</i>	1,59
				CDS_08270	<i>acnA</i>	1,35
				CDS_04910	<i>nuoK</i>	1,35
Nucleotide transport and metabolism	CDS_02460	<i>purB</i>	1,63	CDS_05880		1,94
				CDS_07060	<i>thyX</i>	1,72
	CDS_06330	<i>cmk</i>	1,42	CDS_08980	<i>ndk</i>	1,70
				CDS_06590	<i>purN</i>	1,33
Coenzyme transport and metabolism	CDS_09450	<i>ubiF</i>	1,96	CDS_01680	<i>bioF</i>	2,17
	CDS_05120	<i>thiL</i>	1,81	CDS_04080	<i>folE</i>	1,60
	CDS_02710	<i>nadE</i>	1,64			
	CDS_01790	<i>pdhH</i>	1,56			
	CDS_08040	<i>ubiE</i>	1,52			
	CDS_02030	<i>thiE</i>	1,40			
	CDS_02910	<i>nadD</i>	1,19			
Replication, recombination and DNA repair	CDS_01880		1,87			
	CDS_05050*	<i>recO</i>	1,53			
	CDS_00800	<i>pol III</i>	1,22			
Posttranslational modification, protein turnover, chaperones	CDS_07310	<i>coxW</i>	1,67	CDS_06620	<i>clpB</i>	2,59
	CDS_03060	<i>ccmE</i>	1,41	CDS_07200	<i>grxC2</i>	2,20
	CDS_05090		1,25	CDS_07990		1,71
				CDS_01060	<i>grpE</i>	1,65
				CDS_08630	<i>dsbD</i>	1,46

Carbohydrate,AA,inorganic	CDS_02740	<i>proP</i>	2,01	CDS_07110	<i>glyA</i>	1,48
ion transport and	CDS_05960	<i>pstB</i>	1,73			
metabolism	CDS_05320	<i>carA</i>	1,66			
	CDS_00460	<i>rpe</i>	1,65			
	CDS_08060	<i>lysC</i>	1,30			
	CDS_03380	<i>cutA</i>	1,22			
Lipid transport and	CDS_03460	<i>acpS</i>	1,33			
metabolism	CDS_00950*	<i>ispF</i>	1,12			
Signal transduction	CDS_06840*	<i>ntrX</i>	2,38			
mechanisms	CDS_03390*	<i>virA</i>	1,48			
Transcription				CDS_04040*	<i>rpoH</i>	1,29
				CDS_00730	<i>greA</i>	1,22
Cell wall/ membrane/	CDS_01150	<i>Int</i>	1,51	CDS_07250		1,66
envelope biogenesis				CDS_04980	<i>mraW</i>	1,45
Intracellular trafficking,				CDS_05410	<i>virB3</i>	2,43
secretion, and vesicular				CDS_03800	<i>lepB</i>	1,55
transport				CDS_05390	<i>virB6a</i>	1,51
Map1related protein	CDS_09130	<i>map1-2</i>	2,60			
	CDS_09060	<i>map1-9</i>	2,55			
	CDS_09050	<i>map1-10</i>	2,47			
	CDS_09080	<i>map1-7</i>	2,07			
	CDS_09110	<i>map1-4</i>	1,98			
	CDS_09090	<i>map1-6</i>	1,60			
	CDS_09030	<i>map1-12</i>	1,57			
	CDS_09070	<i>map1-8</i>	1,56			
	CDS_09010	<i>map1-13</i>	1,34			
Translation, ribosomal	CDS_04360	<i>truA</i>	1,78	CDS_05260	<i>infA</i>	1,59
structure and biogenesis	CDS_08540	<i>hemK</i>	1,57	CDS_01290	<i>rplT</i>	1,51
	CDS_09320	<i>thrS</i>	1,53	CDS_04050		1,26
	CDS_05070	<i>rrmJ</i>	1,45	CDS_05020	<i>RsmE</i>	1,20
	CDS_03360	<i>nifR3</i>	1,43			
	CDS_08260	<i>gatC</i>	1,42			
Secondary metabolites				CDS_00470		2,03
biosynthesis						

Hypothetical membrane protein	CDS_03900	2,11	CDS_03260	2,01
	CDS_07690	1,94	CDS_03690*	1,83
	CDS_07410	1,88	CDS_07920	1,54
	CDS_07940*	1,74	CDS_05890	1,48
	CDS_02790	1,72	CDS_02900	1,18
	CDS_02250*	1,64		
	CDS_02330	1,58		
	CDS_00760	1,54		
	CDS_02240	1,51		
	CDS_04730	1,37		
	CDS_07580	1,33		
	CDS_04990	1,31		
Hypothetical exported protein	CDS_08340	1,70	CDS_05140	1,93
	CDS_03860	1,39	CDS_01380	1,39
	CDS_07400	1,31		
	CDS_03870	1,28		
	CDS_01930	1,27		
	CDS_08440	1,13		
Hypothetical protein	CDS_02340	2,04	CDS_03450	2,93
	CDS_08760*	1,76	CDS_04560	2,73
	CDS_05600	1,74	CDS_09210	2,32
	CDS_06800	1,65	CDS_07190	2,30
	CDS_07820	1,63	CDS_06440	2,18
	CDS_08490	1,59	CDS_04190	2,10
	CDS_07490	1,50	CDS_00430	1,92
	CDS_02510	1,46	CDS_02070	1,71
	CDS_04010	1,34	CDS_04390	1,68
	CDS_01720	1,25	CDS_08200	1,68
	CDS_00410	1,19	CDS_03000	1,62
	CDS_02830	1,11	CDS_06760	1,62
	CDS_08750*	1,09	CDS_05720	1,61
	CDS_03990	0,99	CDS_02360	1,45
	CDS_02350	0,97		

*gene with mutation on attenuated strain, Bold: genes identified as differentially expressed in both Gardel and Senegal strains

References

- Allsopp BA (2010) Natural history of *Ehrlichia ruminantium*. *Veterinary parasitology* **167**: 123-135.
- Barre N, Uilenberg G, Morel PC & Camus E (1987) Danger of introducing heartwater onto the American mainland: potential role of indigenous and exotic Amblyomma ticks. *The Onderstepoort journal of veterinary research* **54**: 405-417.
- Bechah Y, El Karkouri K, Mediannikov O, *et al.* (2010) Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. *Genome Res* **20**: 655-663.
- Collins NE, Liebenberg J, de Villiers EP, *et al.* (2005) The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc Natl Acad Sci U S A* **102**: 838-843.
- Didier G, Brezellec P, Remy E & Henaut A (2002) GeneANOVA--gene expression analysis of variance. *Bioinformatics* **18**: 490-491.
- Emboule L, Daigle F, Meyer DF, *et al.* (2009) Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol Biol* **10**: 111.
- Frutos R, Viari A, Ferraz C, *et al.* (2006) Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *Journal of bacteriology* **188**: 2533-2542.
- Martinez D, Vachieri N, Stachurski F, Kandassamy Y, Raliniaina M, Aprelon R & Gueye A (2004) Nested PCR for detection and genotyping of *Ehrlichia ruminantium*: use in genetic diversity analysis. *Annals of the New York Academy of Sciences* **1026**: 106-113.

- Merhej V & Raoult D (2011) Rickettsial evolution in the light of comparative genomics. *Biol Rev Camb Philos Soc* **86**: 379-405.
- Nakanishi N, Abe H, Ogura Y, *et al.* (2006) ppGpp with DksA controls gene expression in the locus of enterocyte effacement (LEE) pathogenicity island of enterohaemorrhagic *Escherichia coli* through activation of two virulence regulatory genes. *Molecular microbiology* **61**: 194-205.
- Nicholson TL, Olinger L, Chong K, Schoolnik G & Stephens RS (2003) Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. *Journal of bacteriology* **185**: 3179-3189.
- Pilet H, Vachieri N, Berrich M, *et al.* (2012) A new typing technique for the Rickettsiales *Ehrlichia ruminantium*: multiple-locus variable number tandem repeat analysis. *Journal of microbiological methods* **88**: 205-211.
- Postigo M, Taoufik A, Bell-Sakyi L, Bekker CP, de Vries E, Morrison WI & Jongejan F (2008) Host cell-specific protein expression in vitro in *Ehrlichia ruminantium*. *Veterinary microbiology* **128**: 136-147.
- Pruneau L, Emboule L, Gely P, *et al.* (2012) Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol Med Microbiol* **64**: 66-73.
- Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* **33** (Web Server issue): W116-W120
- Raliniaina M, Meyer DF, Pinarello V, *et al.* (2010) Mining the genetic diversity of *Ehrlichia ruminantium* using map genes family. *Vet Parasitol* **167**: 187-195.
- Renesto P, Rovery C, Schrenzel J, *et al.* (2008) *Rickettsia conorii* transcriptional response within inoculation eschar. *PLoS One* **3**: e3681.
- Rouillard, J.M., Zuker, M. and Gulari, E. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res*, **31**, 3057-3062.

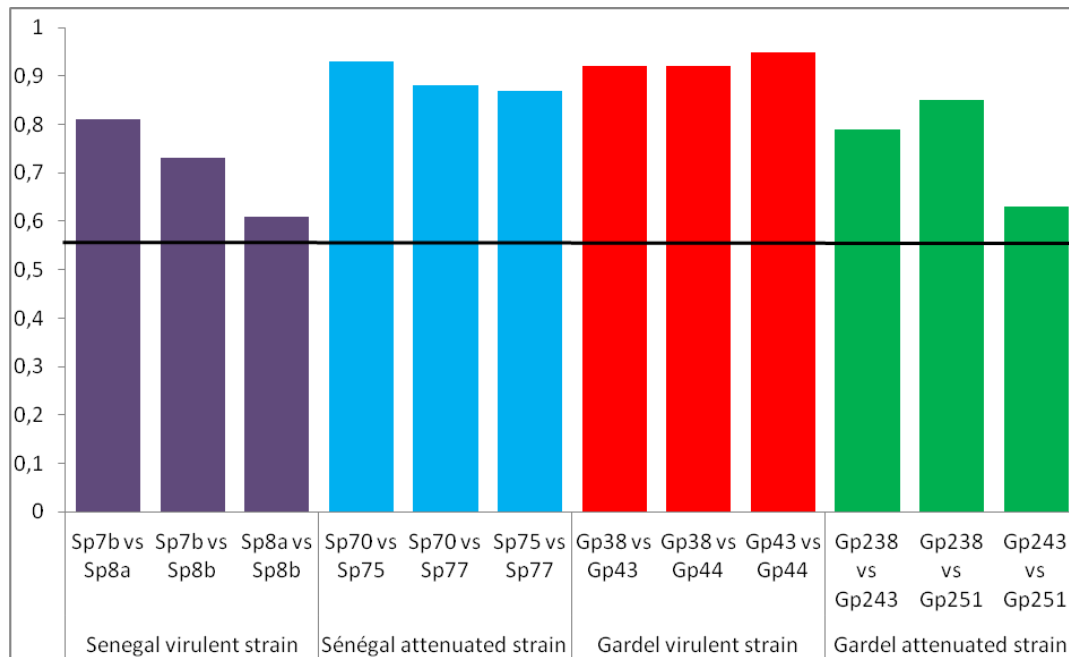


Figure 1-A: linear correlation between the different replicates

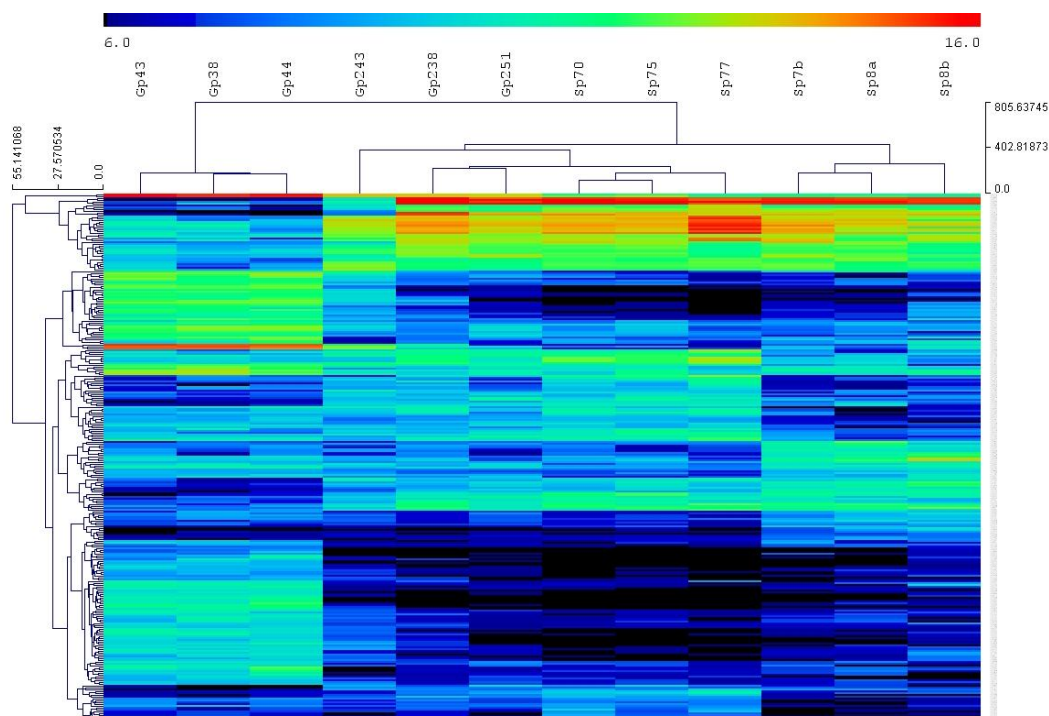


Figure 1-B: clustering of biological replicates with the 300 best probes

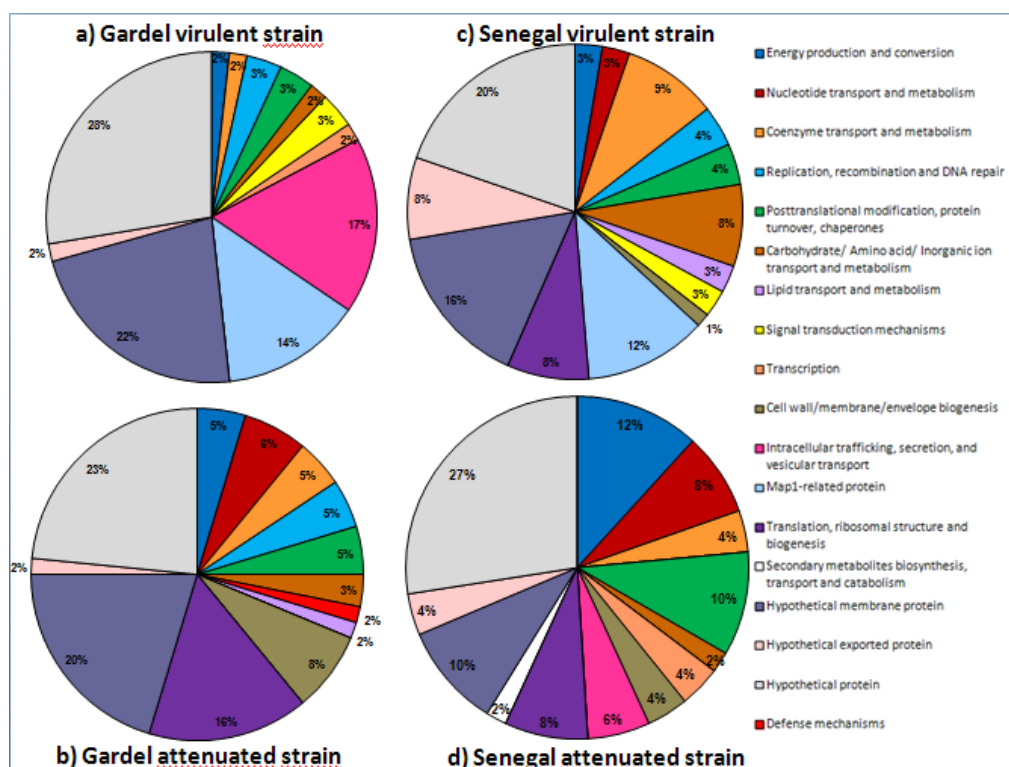


Figure 2: Proportion of COGs of genes over-expressed for virulent (a) and attenuated (b) Gardel, and virulent (c) and attenuated Senegal (d).

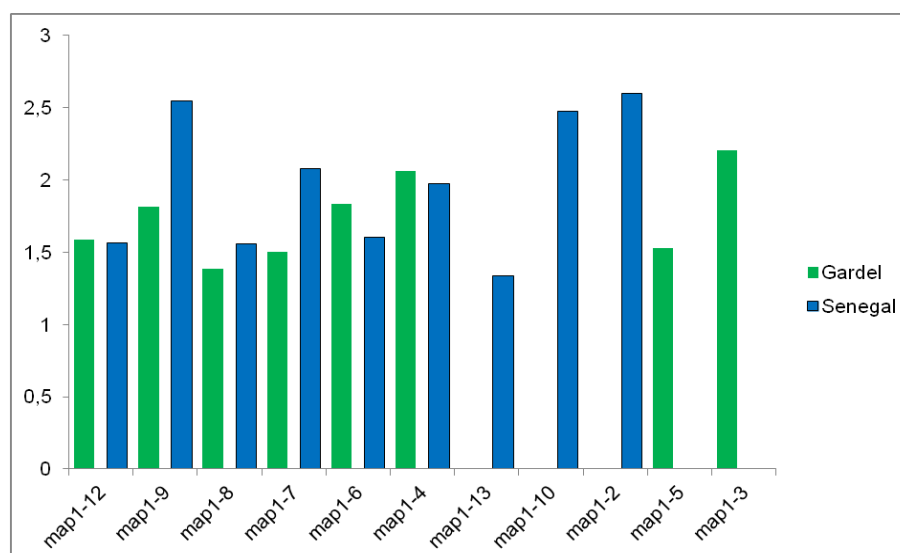


Figure 3: Overexpression of *map1*-related genes for virulent Gardel and Senegal strains. FC-log2 values are represented for two strains by bar charts.

IV. Chapitre 4:

Discussion générale

SOMMAIRE

1.	Nature des souches Gardel et Sénégal	156
2.	Mécanismes de virulence et d'atténuation d' <i>E. ruminantium</i>	159
2.1.	Développement intracellulaire: expression différentielle des gènes du métabolisme	159
2.2.	Echappement aux mécanismes de défense de l'hôte	161
2.3.	Déterminants de virulence.....	163
2.4.	Echappement à la réponse immunitaire de l'hôte	164
2.5.	Mécanisme d'atténuation lié aux mutations génomiques	165

1. Nature des souches Gardel et Sénégal

Au sein de l'espèce *E. ruminantium*, il existe une très grande diversité génétique des souches quelque soit la zone géographique étudiée ([Allsopp, 2010](#)). Les premiers essais vaccinaux effectués ont montré leur limite lors d'infection avec des souches très différentes de la souche vaccinale. L'élaboration de nouveaux vaccins nécessite de cibler des déterminants de virulence communes à l'ensemble des souches ou alors d'utiliser des cocktails de souches vaccinales (vaccin inactivés ou atténués) les mieux adaptés à chaque zone géographique. Dans ce contexte, nous avons entrepris l'analyse du transcriptome de deux souches d'*E. ruminantium* : la souche Gardel isolée en Guadeloupe et la souche Sénégal isolée au Sénégal. Le choix de ces deux souches a été guidé par le fait qu'elles sont phylogénétiquement très éloignées ([Raliniaina, et al., 2010](#), [Pilet, et al., 2012](#)), elles ne protègent que très peu l'une contre l'autre et nous avons à disposition les souches atténuées correspondantes sur lesquelles nous avons également mené une étude du transcriptome. De plus, les souches Gardel et Sénégal s'atténuent à des vitesses différentes ce qui suggère un mécanisme d'atténuation différente.

Afin d'obtenir des résultats robustes du transcriptome d'*E. ruminantium*, et d'éviter une seconde validation par q-RTPCR, trois réplicats biologiques pour chaque souche à chaque conditions expérimentales ont été hybridés. La première difficulté était d'obtenir des cultures synchrones. Pour Gardel virulente et Sénégal atténuée, les courbes de croissance des réplicats étaient globalement synchrones. Pour Gardel atténuée, même si la courbe de croissance était similaire entre réplicats, les quantités de bactéries étaient différentes pour les temps précoces d'infection.

La souche Sénégal s'atténuant rapidement en culture, nous avons du travailler avec des bactéries directement isolées de l'animal et donc n'ayant jamais été cultivées *in vitro*. L'adaptation de Sénégal virulente à la culture *in vitro* a été très difficile, avec des cycles infectieux extrêmement longs (lyse des cellules hôtes observée entre 15 et 30 jours après infection). Le challenge

était d'obtenir des réplicats biologiques avec des cycles infectieux identiques. Après de nombreux essais, nous avons pu obtenir trois réplicats biologiques issus de différents isollements, présentant un profil de courbes de croissance similaire avec une longue phase de latence et une lyse totale des cellules hôtes infectées au bout de 144 hpi. Cependant, d'autres essais de culture *in vitro* devront être menés afin de pouvoir déterminer les conditions optimales de culture de la souche Sénégal virulente.

Au vue des résultats obtenus pour les 4 souches, la *Rickettsiales E. ruminantium* semble montrer une adaptation progressive (diminution de la durée de la phase de latence) lorsque celle-ci est cultivée *in vitro* en cellules endothéliales. Cette adaptation semble être également corrélée à l'atténuation puisque le cycle infectieux des souches atténuées est plus court que ceux des souches virulentes.

De plus, en s'appuyant sur nos observations microscopiques réalisées tout au long de l'infection, nous avons pu constater que le développement *in vitro* des 4 souches n'était pas identique en termes d'apparition, de morphologie et d'organisation des morulas. Il a été montré que les CE de Gardel atténuée étaient plus nombreux à la surface des cellules lors de l'infection par rapport à la souche virulente par observation au microscope électronique (données non montrées), ce qui explique le plus grand nombre de morulas dans les cellules infectées par la souche atténuée. Le fait que la souche Sénégal forme des morulas en grappes de raisin alors que Gardel virulente a plutôt quelques grosses morulas, suggèrent que l'adhésion aux cellules et l'infection est différente entre les 2 souches. En se basant sur ces observations et les allures des courbes de croissance, nous avons déterminé trois temps post-infection différents entre les souches qui devraient correspondre aux mêmes stades de développement. Cependant, bien qu'au temps intermédiaire T2, *E. ruminantium* soit majoritairement sous la forme CR, il n'est pas à exclure un mélange des différents stades de développement de la bactérie. De plus, le choix des temps post infection T1 et T2 est difficile à faire étant donné la différence de comportement entre les

souches Gardel et Sénégal et entre les souches atténuées et virulentes. Un indicateur de stade de développement tel qu'un gène ou plusieurs gènes spécifiques d'un stade serait nécessaire de manière à corrélérer un temps post infection à un stade donné.

L'enrichissement en transcrits procaryotes des différents échantillons générés, a ensuite été effectué à l'aide de la méthode SCOTS. Cette méthode a été validée chez les souches Gardel ([Emboule, et al., 2009](#)) et au cours de cette thèse, nous l'avons adapté avec succès aux souches Sénégal. En effet, comme attendu, nous avons observé une diminution des transcrits ARN ribosomiaux au cours des captures pour les souches Sénégal, mais cette diminution était beaucoup moins prononcée que ce qui a pu être observé chez Gardel. Il est possible que cette méthode de capture sélective entraîne des biais pour la détermination des gènes différentiellement transmis. Nous avons fait le choix d'hybrider sur les lames les échantillons sans captures et les échantillons capturés avec 2 rounds SCOTS pour les temps intermédiaires (T2) et tardifs (T3). Les résultats ainsi obtenus en microarrays, nous permettront de vérifier les biais relatifs à SCOTS pour chacune des souches et pour les différents stades de développement. D'autres techniques existent pour la sélection de transcrits de bactéries intracellulaires comme l'utilisation des 2 kits développés par la société Invitrogen : le kit MICROBEnrich permettant d'éliminer l'ARN eucaryote couplé au kit MICROBExpress qui permet d'éliminer l'ARN ribosomal procaryote. Mais l'utilisation de ces kits nécessite des quantités d'ARN de départ beaucoup plus importante (5 µg) et des coûts élevés, alors que la méthode SCOTS elle, nécessite dix fois moins de matériel de départ avec un coût moindre.

2. Mécanismes de virulence et d'atténuation d'*E. ruminantium*

Afin d'identifier les mécanismes de virulence et d'atténuation de la *Rickettsiales E. ruminantium*, nous avons entrepris une étude du transcriptome globale des 4 souches Gardel et Sénégal virulente et atténuée. Les résultats d'hybridation obtenus pour les différents réplicats d'une même souche, montraient une bonne homogénéité avec un coefficient de corrélation linéaire supérieur à 0.6. Cette homogénéité renforce nos données obtenues en microarrays, nous évitant ainsi de faire une validation du niveau d'expression des gènes par qPCR.

Nous avons dans un premier temps mené une analyse du transcriptome au cours du développement de la souche Gardel virulente, puis dans un second temps, une analyse du transcriptome des souches virulentes *vs* atténuées au stade infectieux (T3). Ces analyses ont révélé l'expression différentielle de gènes impliqués dans l'invasion, le développement intracellulaire, l'échappement aux mécanismes de défense de l'hôte. Nos résultats suggèrent également un important processus de modification de la membrane. Pour Gardel et Sénégal, en comparant les souches virulentes aux atténués, il y a une modification de l'expression des protéines de membrane, pour des protéines clés dans la réponse immunitaire telles que les MAP ou pour des protéines membranaires de fonction inconnue et qui pourraient être des candidats vaccinaux intéressants. Enfin, nous avons pu mettre en évidence des phénomènes de mutation chez Sénégal atténuée et qui sont corrélés à la diminution d'expression de certains gènes alors que chez Gardel la présence de ces mutations est rare.

2.1. Développement intracellulaire: expression différentielle des gènes du métabolisme

L'une des fonctions clés du métabolisme est la production d'énergie. Celle ci est essentielle pour l'élaboration des nombreux processus nécessaires au cycle de développement. La sur-expression des gènes

impliqués dans la production d'énergie était détectée pour les 4 souches mais avec une proportion plus élevée pour les souches atténuées. Pour la souche Gardel virulente, cette fonction semblait être plus importante au stade CR intracellulaire (T2) que pour le stade extracellulaire CE (article 1). Les gènes codant pour les différentes chaînes de l'ATP synthase tels qu'*atpC*, *atpE* et *atpG*, étaient sur-exprimés pour les souches atténuées. Ces résultats suggèrent que la voie de production d'ATP par phosphorylation oxydative est plus active chez les souches atténuées.

De plus, les gènes impliqués dans le transport et le métabolisme des nucléotides étaient aussi sur-exprimés chez les souches atténuées. Un de ces gènes, *ndk* code pour la Nucleoside diphosphate kinase qui catalyse la formation de nucléoside triphosphate nécessaire à la synthèse d'ADN, à partir de nucléoside diphosphate et d'ATP. Cette activation privilégiée de la voie de production d'ATP nécessaire d'une part pour l'énergie mais aussi pour la synthèse d'ADN, pourrait contribuer au cycle infectieux plus court des souches atténuées dû à une répllication de l'ADN commençant dès l'infection de nouvelles cellules hôtes. Chez la souche Sénégal virulente, une surexpression du gène *sdhC* qui catalyse l'oxydation du succinate en fumarate était observée. Chez *Salmonella enterica* serovar Typhimurium, la mutation de *sdhCDA* aboutissait à l'atténuation de la bactérie ([Mercado-Lubo, et al., 2008](#)). Il se peut que le gène *sdhC*, soit impliqué dans les mécanismes d'atténuation d'*E. ruminantium* comme c'est le cas pour *S. typhimurium*.

La sur-expression importante de gènes impliqués dans le transport et le métabolisme de coenzymes, de carbohydrates, d'acides aminés, et d'ions inorganiques observée chez la souche Sénégal virulente est intéressante. Ce résultat peut être dû à une utilisation moindre des éléments de la cellule hôte et pourrait traduire l'adaptation moins bonne de cette souche au système de culture *in vitro*.

Nos résultats ont montré également, une sur-expression des gènes impliqués dans l'acquisition et le transport du fer (*fdxA* et *fdxB*) chez les 2

souches atténuées par rapport aux souches virulentes. Cependant, la souche Gardel virulente sur-exprimait aussi un élément d'acquisition du fer notamment une ferrochélatase, *hemA*. Le fer est un élément essentiel pour la plupart des organismes vivants, y compris pour les bactéries, et la déplétion de celui-ci peut induire une inhibition de l'activité enzymatique ([Paradkar, et al., 2008](#)).

Les gènes impliqués dans la traduction, la biogenèse et la structure des ribosomes représentaient une proportion importante de gènes sur-exprimés chez les souches Gardel et Sénégal atténuées et chez la souche Sénégal virulente. Ce résultat semble indiquer que pour ces 3 souches, il y ait une production intense de protéine qui pourrait être lié à la production d'enzymes intervenant dans les fonctions métaboliques telles que la production d'enzyme ou encore le transport et le métabolisme d'éléments (coenzymes, acides aminés...). Pour la souche Gardel virulente, comparé à la souche atténuée au stade T3, il n'y avait pas de gènes sur-exprimés impliqués dans la traduction et la biogenèse des ribosomes. De plus, la comparaison du transcriptome global au cours du développement chez Gardel virulente, indiquait une proportion de gènes impliqués dans ces fonctions, plus importante au stade T2 comparé au stade T3. La souche Gardel virulente semble donc avoir une importante activité de production protéique au stade CR plutôt qu'au stade CE.

2.2. Echappement aux mécanismes de défense de l'hôte

A l'intérieur de sa cellule hôte, *E. ruminantium* est confrontée à des variations de température, de pH et du statut oxydo-réducteur lié à la production d'espèces oxygénées actives.

Chez les bactéries intracellulaires, les thiorédoxines (Trx) et glutarédoxines (Grx) agissent comme des anti-oxydants en facilitant la réduction de protéines oxydées ([Holmgren, 1989](#)). La délétion de gènes appartenant au complexe Grx est corrélée à une diminution de la résistance au stress

oxydatif, ce qui montre leur implication dans divers processus de survie ([Bjur, et al., 2006](#)). Les gènes codant pour des Trx et Grx et d'autres gènes impliqués dans la lutte contre le stress oxydatif étaient exclusivement surexprimés chez les souches atténuées Gardel et Sénégal indiquant donc une meilleure adaptation des souches atténuées à ce type de stress. Le gène hypothétique ERGA_CDS_07190 code pour une protéine de la famille des BolA like proteins qui sont induits en cas de stress. Ce gène était surexprimé par la souche Sénégal atténuée et est localisée sur le même opéron que *grxC2* (CDS_07200), il semblerait donc que ces 2 gènes soient activés ensemble pour lutter contre le stress oxydatif. Les bactéries intracellulaires doivent également faire face au stress osmotique qui est lié à la variation de pH entre leur cytoplasme et le milieu extérieur. Pour lutter contre ce stress, elles possèdent des systèmes de transport et/ou de synthèse d'osmoprotectants ou solutés compatibles qui assurent le maintien de l'homéostasie par leur accumulation dans le cytoplasme ou leur rejet selon le type de stress osmotique ([Smith, 1996](#)). La comparaison du transcriptome des souches Gardel virulente et atténuée de même que la comparaison du transcriptome de la souche virulente au cours du développement, ne nous a pas permis d'identifier des gènes différentiellement exprimés et impliqués dans la lutte contre le stress osmotique. Par contre, pour les souches Sénégal, nous avons mis en évidence 6 gènes différentiellement exprimés et impliqués dans la lutte contre le stress osmotique incluant les gènes *nuoE*, *nuoH* et *nuoK* qui codent pour les chaînes d'oxydoréductase du NADH-quinone E, H et K respectivement. Ces gènes codent pour des protéines impliqués dans la production d'énergie mais pouvant aussi contribuer au maintien de l'homéostasie comme cela a été observé chez *R. conorii* ([Renesto, et al., 2008](#)).

2.3. Déterminants de virulence

Les bactéries pathogènes mettent en œuvre différentes stratégies pour infecter les cellules eucaryotes. L'une de ces stratégies, fait intervenir la sécrétion de facteurs qui facilitent l'adhésion, l'internalisation ou la virulence. Ces facteurs sont sécrétés par les systèmes de sécrétion tels que le SST4. Le SST4 est connu pour être impliqué dans la pathogenèse de plusieurs bactéries en translocant des molécules telles que les facteurs de virulence dans les cellules hôtes eucaryotes. Les gènes codant pour le SST4 étaient sur-exprimés au stade CE pour la souche Gardel virulente. Le SST4 pourrait jouer ainsi un rôle lors de l'infection de la cellule hôte par *E. ruminantium*.

Le gène *dksA* était à la fois sur-exprimé au stade CE comparé au stade CR mais aussi sur-exprimé chez la souche Gardel virulente comparée à la souche Gardel atténuée. *DksA* est impliqué dans la régulation de gènes de virulence chez *S. typhimurium* et *E. coli* Entéro-Hémorragique (EHEC) ([Nakanishi, et al., 2006](#)). A la lyse cellulaire, *E. ruminantium* via *dksA* pourrait réguler des gènes de virulence, lui permettant ainsi de ré-infecter de nouvelles cellules hôtes.

Par la recherche de motifs protéiques, nous avons pu identifier deux gènes hypothétiques avec des motifs eucaryotes importants pour les interactions avec les acides nucléiques ou les protéines de l'hôte. Il s'agit du gène ERGA_CDS_06440 sur-exprimé par Sénégal atténué et du gène ERGA_CDS_08730 sur-exprimé par la souche Gardel virulente. ERGA_CDS_06440 présentant un motif ankyrine pourrait agir sur la modulation transcriptionnelle des gènes de la cellule hôte comme cela a été montré pour *ankA* chez *A. phagocytophilum* ([Al-Khodor, et al., 2010](#)). La fonction de ces protéines atypiques chez les bactéries est encore largement inconnue mais elles sont fortement suspectées d'être impliquées dans le cycle cellulaire, l'organisation du cytosquelette, la

toxicité ou encore la réponse inflammatoire ([Wakeel, et al., 2010](#)). Le gène hypothétique ERGA_CDS_00640 était sur-exprimé au stade CE comparé au stade CR. Ce gène était aussi fortement exprimé chez les 4 souches mais avec un niveau équivalent. Dans une récente étude menée sur *A. phagocytophilum*, l'orthologue de ce gène, APH_1235 a été montré comme étant impliqué dans l'infection des cellules hôtes. En effet, le blocage de ce gène par des anticorps réduisait le taux d'infection des cellules hôtes ([Mastronunzio, et al., 2012](#)).

2.4. Echappement à la réponse immunitaire de l'hôte

La comparaison du transcriptome globale entre les souches virulentes et atténuées pour Gardel et Sénégal a révélé une sur-expression des gènes de la famille multi-génique *map*, exclusivement chez les souches virulentes. Ainsi, 8 et 9 gènes *map* étaient sur-exprimés chez Gardel et Sénégal respectivement, dont 6 en commun entre les deux souches. Une étude de l'expression des protéines MAP dans les cellules de l'hôte comparée à l'expression dans les cellules de tiques a montré une expression spécifique de certaines MAP vis-à-vis du type cellulaire ([Postigo, et al., 2008](#)). Il existe aussi une différence d'expression des gènes *map1* et *map1-1* dans les tiques, intestin ou glande salivaire en fonction du gorgement. Les protéines MAP sont probablement impliquées dans les interactions ou les adaptations d'*E. ruminantium* à son hôte et son vecteur. MAP1 induit une réponse humorale importante chez l'animal lors de l'infection mais il a été montré que les anticorps anti-MAP1 ne protégeaient pas lors d'un transfert d'anticorps de l'animal immun vers l'animal naïf. Le fait d'une sur-expression exclusive de ces protéines chez les souches virulentes suggèrent qu'elles puissent jouer le rôle de leurre du système immunitaire entraînant une production d'anticorps et une réponse dirigée contre elles et non dirigée contre d'autres protéines de membrane qui pourraient induire une protection contre *E. ruminantium*. Chez les souches atténuées, des protéines hypothétiques de

membrane sont sur-exprimées à la surface des CE, indiquant une modification importante de la nature de la membrane. Ces résultats suggèrent que cette modification induit une réponse immunitaire protectrice chez les souches atténuées ce qui n'est pas le cas chez les souches virulentes. Les protéines de membranes hypothétiques qui sont surexprimées pour les souches atténuées peuvent être des candidats vaccinaux d'intérêt. Ces protéines sont polymorphes vis-à-vis des souches d'*E. ruminantium* ce qui renforcent l'hypothèse que les antigènes protecteurs sont portés sur des protéines polymorphes empêchant alors une bonne protection entre les souches différentes.

2.5. Mécanisme d'atténuation lié aux mutations génomiques

A l'intérieur de la cellule hôte, l'ADN génomique bactérien peut être endommagé ce qui nécessite des mécanismes de réparation de l'ADN lors de la réplication. Des gènes impliqués dans ces mécanismes étaient sur-exprimés chez les souches Gardel et Sénégal virulente et chez Gardel atténuée avec des proportions équivalentes. Étonnamment, il n'y avait pas de gènes de cette catégorie fonctionnelle chez la souche Sénégal atténuée. De plus, le gène *recO*, sur-exprimé chez Sénégal virulente était tronqué chez Sénégal atténuée. Le gène *recO* code pour une protéine du complexe RecFOR qui est impliqué dans la réparation de l'ADN ([Chow & Courcelle, 2004](#)). La mutation de ce gène pourrait être à l'origine de mutations d'autres gènes. En effet, parmi les gènes différentiellement exprimés entre Sénégal virulente et atténuée, 9 gènes étaient trouvés tronqués ou cassés chez Sénégal atténuée. La mutation du gène *recO* entraînant la mutation d'autres gènes pourrait expliquer l'atténuation plus rapide de la souche Sénégal. Dans une récente étude menée chez *R. prowazekii*, les auteurs ont montré que l'inactivation du gène *recO* entraînait l'atténuation des souches ([Bechah, et al., 2010](#)).

D'autres gènes impliqués dans la réplication et la réparation de l'ADN tels que *nth* codant pour l'endonuclease III et ERGA_CDS_05820 codant pour une Holliday junction resolvase, étaient sur-exprimés chez la souche Gardel atténuée. Cela pourrait expliquer l'atténuation plus lente de la souche Gardel ainsi qu'une proportion plus faible de mutations entre les génomes de Gardel virulente et atténuée.

V. Chapitre 5 :

Conclusion et perspectives

L'objectif de ce travail de thèse était de pouvoir mettre en évidence les gènes impliqués dans les mécanismes de virulence et d'atténuation de la *Rickettsiales E. ruminantium*. Pour pouvoir identifier ces gènes, nous avons mené une étude du transcriptome sur deux souches d'origines géographiques différentes. Il s'agit des souches, Gardel isolée en Guadeloupe et Sénégal isolé au Sénégal. Ces souches phylogénétiquement différentes ont pu être atténuées *in vitro* par passages successives en cellules endothéliales. De plus, ces souches ont un développement *in vitro* et un mécanisme d'atténuation différents.

A partir des données microarrays générées lors de la précédente thèse au sein de l'équipe, nous avons faite une première analyse qui reposait sur la comparaison du transcriptome globale de la souche Gardel virulente au cours du développement. Les résultats obtenus permettent de mettre en évidence une activité métabolique intense au stade corps réticulé qui est, le stade intracellulaire non infectieux de la bactérie. Ce résultat est concordant au fait qu'à ce stade de développement, *E. ruminantium* est dans sa phase exponentielle de croissance. Elle active donc des gènes lui permettant de se multiplier dans les cellules endothéliales. Au stade corps élémentaire qui correspond à la forme libre et infectieuse de la bactérie, nous avons identifié un facteur de transcription de gènes de virulence et des gènes impliqués dans les mécanismes de transduction de signaux. On peut penser qu'à ce stade, *E. ruminantium* active un ensemble de gènes lui permettant d'infecter de nouvelles cellules hôtes.

Au cours de cette thèse, nous avons généré de nouvelles données microarrays plus robustes aux différents stades de développement pour les 4 souches. Ces données seront bientôt analysées en comparant le niveau d'expression des gènes au cours de l'infection. Elles permettront d'une part de confirmer les données obtenues pour la souche Gardel virulente, et d'autre part d'identifier les gènes impliqués dans le développement des 3 autres souches Gardel atténuée et Sénégal virulente et atténuée. De plus, les données obtenues pour le temps précoce sont exploitables et seront

comparées aux données obtenues pour le temps intermédiaires permettant ainsi d'identifier les gènes différentiellement exprimés entre ces 2 stades de développement ce qui n'a jamais été fait jusqu'à présent.

La comparaison des données entre les 4 souches aux différents temps permettra d'identifier des gènes spécifiques de chaque stade de développement et qui pourront être considérés comme marqueurs de stade de développement. Ils seront utilisés pour choisir les stades post-infection à comparer entre les souches.

Dans un second temps, les comparaisons du transcriptome des 4 souches au stade CE, nous a permis d'émettre l'hypothèse d'un mécanisme de virulence lié à l'échappement à la réponse de l'hôte. En effet, les souches virulentes échapperaient *in vivo* à la reconnaissance par le système immunitaire en exprimant à leur surface les protéines membranaires de la famille MAP induisant ainsi une réponse inefficace et non protective. Les souches atténuées, elles exprimeraient à leur surface d'autres protéines membranaires qui *in vivo* entraîneraient une reconnaissance du pathogène par le système immunitaire et donc une élimination plus précoce des bactéries atténuées. Des études sur la fonction des MAP et sur la fonction de ces protéines de membranes hypothétiques et de leur efficacité comme candidats vaccinaux seront menés pour confirmer cette hypothèse. Notamment, il est envisagé de vérifier la capacité de séroneutralisation des anticorps antiMap sur la pénétration des bactéries dans la cellule hôte à la fois chez la souche virulente et atténuée pour voir quel est l'impact des MAP dans la pénétration dans la cellule.

L'étude du protéome de membrane de ces 4 souches d'*E. ruminantium* est actuellement en cours. Les résultats obtenus seront comparés aux résultats de microarrays nous permettant ainsi de valider l'expression des gènes au niveau protéique.

L'analyse du génome des 4 souches est actuellement en cours et permettra d'identifier les gènes mutés entre les souches virulentes et atténuées qui peuvent potentiellement être impliqués dans l'atténuation. Nous avons pu

identifier pour la souche Sénégal, des mutations directement liés à la sous expression des gènes chez l'atténuee par rapport à la virulente. Dans le cas de Gardel, la corrélation n'est pas aussi évidente et la modification d'expression des gènes est probablement liée à des phénomènes plus fins de régulation notamment des mutations au niveau des opérons ou des modifications post-traductionnelles. Des investigations plus poussées sur le différentiel d'expression des gènes et les modifications au niveau des génomes sont en cours.

La méthode sélective de captures des transcrits (SCOTS) est nécessaire pour l'analyse des stades précoce et intermédiaire. Pour le stade tardif, les résultats obtenus sur les échantillons sans captures permettaient une bonne détection des transcrits sans de biais liés à la méthode SCOTS. Les dernières avancées en biologie moléculaire permettent aujourd'hui de séquencer directement les ARNs issus d'un échantillon. L'utilisation de cette méthode permettrait de limiter les biais liés à la manipulation, aux différentes méthodes de captures sélectives et aux microarrays. De plus, elle permettrait de mettre en évidence les ARNs non codants qui peuvent être impliqués dans la régulation des gènes et d'ARNm.

L'utilisation de la mutagenèse dirigée reste encore difficile à mettre en oeuvre pour les bactéries intracellulaires obligatoires, des essais peuvent être entrepris sur *E. ruminantium*, offrant ainsi des possibilités de validation fonctionnelle de gènes d'intérêts. Beaucoup de gènes différentiellement exprimés, ont des fonctions hypothétiques. La recherche de l'impact de ces gènes dans le développement et la virulence de la bactérie, par mutagenèse dirigée est nécessaire.

Une autre méthode de validation des gènes d'intérêt envisageable, est la complémentation dans l'hôte hétérologue *Bartonella henselae*. Ce modèle semble être bien adapté pour tester la fonctionnalité de certains gènes conservés d'*E. ruminantium*.

Nos résultats ont révélé également des mécanismes de lutte mise en place par *E. ruminantium* pour lutter contre les stress oxydatif et osmotique. Des

essais *in vitro* tels que le suivi de la concentration en espèces oxygénés activés ou l'augmentation de l'osmolarité du milieu de culture, peuvent être envisageable afin de vérifier la réponse des souches vis-à-vis de ces stress.

Les transporteurs et systèmes d'acquisition de métabolites semblent être importants pour *E. ruminantium*. Des expériences (effets doses) d'inhibition des éléments transportés tels que le fer, devront être réalisées pour confirmer leur importance dans la survie et la virulence intracellulaire de la bactérie.

Une étude du protéome globale des 4 souches d'*E. ruminantium* est actuellement en cours. Les résultats obtenus seront comparés aux résultats obtenus au cours de cette thèse et aux données obtenues sur la comparaison des génomes de ces 4 souches, permettant ainsi une vision globale du gène à la protéine.

Des études complémentaires du transcriptome et du protéome de l'hôte et du vecteur infectés avec *E. ruminantium* doivent être envisagées afin de pouvoir clairement identifier les gènes intervenant dans les mécanismes et interactions du pathogène avec son hôte et son vecteur.

VI. Références bibliographiques

Adakal H, Stachurski F, Konkobo M, *et al.* (2010) Efficiency of inactivated vaccines against heartwater in Burkina Faso: Impact of *Ehrlichia ruminantium* genetic diversity. *Vaccine*.

Al-Khodor S, Price CT, Kalia A & Abu Kwaik Y (2010) Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol* **18**: 132-139.

Allsopp BA (2010) Natural history of *Ehrlichia ruminantium*. *Veterinary parasitology* **167**: 123-135.

Allsopp MT, Louw M & Meyer EC (2005) *Ehrlichia ruminantium*: an emerging human pathogen? *Ann N Y Acad Sci* **1063**: 358-360.

Barre N, Uilenberg G, Morel PC & Camus E (1987) Danger of introducing heartwater onto the American mainland: potential role of indigenous and exotic *Amblyomma* ticks. *The Onderstepoort journal of veterinary research* **54**: 405-417.

Bechah Y, El Karkouri K, Mediannikov O, *et al.* (2010) Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. *Genome Res* **20**: 655-663.

Bekker CP, Postigo M, Taoufik A, Bell-Sakyi L, Ferraz C, Martinez D & Jongejan F (2005) Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. *J Bacteriol* **187**: 4782-4791.

Bezuidenhout JD (1987) Natural transmission of heartwater. *Onderstepoort J Vet Res* **54**: 349-351.

Bjur E, Eriksson-Ygberg S, Aslund F & Rhen M (2006) Thioredoxin 1 promotes intracellular replication and virulence of *Salmonella enterica* serovar Typhimurium. *Infect Immun* **74**: 5140-5151.

Camus E & Barre N (1995) Vector situation of tick-borne diseases in the Caribbean Islands. *Vet Parasitol* **57**: 167-176.

Camus E, Maillard JC, Ruff G, Pepin L, Naves M & Matheron G (1996) Genetic resistance of Creole goats to cowdriosis in Guadeloupe. Status in 1995. *Annals of the New York Academy of Sciences* **791**: 46-53.

Chow KH & Courcelle J (2004) RecO acts with RecF and RecR to protect and maintain replication forks blocked by UV-induced DNA damage in *Escherichia coli*. *The Journal of biological chemistry* **279**: 3492-3496.

Collins NE, Liebenberg J, de Villiers EP, *et al.* (2005) The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc Natl Acad Sci U S A* **102**: 838-843.

Cowdry EV (1925) Studies on the Etiology of Heartwater : Ii. *Rickettsia Ruminantium* (N. Sp.) in the Tissues of Ticks Transmitting the Disease. *The Journal of experimental medicine* **42**: 253-274.

Daigle F, Graham JE & Curtiss R, 3rd (2001) Identification of *Salmonella typhi* genes expressed within macrophages by selective capture of transcribed sequences (SCOTS). *Molecular microbiology* **41**: 1211-1222.

Didier G, Brezellec P, Remy E & Henaut A (2002) GeneANOVA-gene expression analysis of variance. *Bioinformatics* **18**: 490-491.

Du Plessis JL (1985) A method for determining the *Cowdria ruminantium* infection rate of *Amblyomma hebraeum*: effects in mice injected with tick homogenates. *Onderstepoort J Vet Res* **52**: 55-61.

Du Plessis JL & Kumm NA (1971) The passage of *Cowdria ruminantium* in mice. *Journal of the South African Veterinary Medical Association* **42**: 217-221.

Dumler JS, Barbet AF, Bekker CP, *et al.* (2001) Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol* **51**: 2145-2165.

Dunning Hotopp JC, Lin M, Madupu R, *et al.* (2006) Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* **2**: e21.

Emboule L, Daigle F, Meyer DF, *et al.* (2009) Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol Biol* **10**: 111.

Euzeby JP & Tindall BJ (2001) Nomenclatural type of orders: corrections necessary according to Rules 15 and 21a of the Bacteriological Code (1990 Revision), and designation of appropriate nomenclatural types of classes and subclasses. Request for an opinion. *Int J Syst Evol Microbiol* **51**: 725-727.

Faburay B, Geysen D, Ceesay A, *et al.* (2007) Immunisation of sheep against heartwater in The Gambia using inactivated and attenuated *Ehrlichia ruminantium* vaccines. *Vaccine* **25**: 7939-7947.

Faucher SP (2007) Identification et caractérisation de gènes impliqués dans la virulence de *Salmonella typhi* suite à une analyse globale par biopuces de l'infection de macrophages humains en culture. Thesis, Université de Montréal, Montréal.

Frutos R, Viari A, Vachiery N, Boyer F & Martinez D (2007) *Ehrlichia ruminantium*: genomic and evolutionary features. *Trends Parasitol* **23**: 414-419.

Frutos R, Viari A, Ferraz C, *et al.* (2006) Comparative genomics of three strains of *Ehrlichia ruminantium*: a review. *Ann N Y Acad Sci* **1081**: 417-433.

Frutos R, Viari A, Ferraz C, *et al.* (2006) Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *J Bacteriol* **188**: 2533-2542.

Holmgren A (1989) Thioredoxin and glutaredoxin systems. *J Biol Chem* **264**: 13963-13966.

Jongejan F, Vogel SW, Gueye A & Uilenberg G (1993) Vaccination against heartwater using *in vitro* attenuated *Cowdria ruminantium* organisms. *Rev Elev Med Vet Pays Trop* **46**: 223-227.

Jongejan F, Zandbergen TA, van de Wiel PA, de Groot M & Uilenberg G (1991) The tick-borne rickettsia *Cowdria ruminantium* has a Chlamydia-like developmental cycle. *Onderstepoort J Vet Res* **58**: 227-237.

Kelly PJ, Yunker CE, Mason PR, Tagwira M & Matthewman LA (1992) Absence of antibody to *Cowdria ruminantium* in sera from humans exposed to vector ticks. *S Afr Med J* **81**: 578.

Le Brigand K & Barbry P (2007) Mediante: a web-based microarray data manager. *Bioinformatics* **23**: 1304-1306.

Marcelino I, Vachier N, Amaral AI, *et al.* (2007) Effect of the purification process and the storage conditions on the efficacy of an inactivated vaccine against heartwater. *Vaccine* **25**: 4903-4913.

Marcelino I, de Almeida AM, Ventosa M, *et al.* (2012) Tick-borne diseases in cattle: Applications of proteomics to develop new generation vaccines. *J Proteomics*.

Marcelino I, de Almeida AM, Brito C, *et al.* (2012) Proteomic analyses of *Ehrlichia ruminantium* highlight differential expression of MAP1-family proteins. *Vet Microbiol* **156**: 305-314.

Martinez D, Maillard JC, Coisne S, Sheikboudou C & Bensaid A (1994) Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium*. *Vet Immunol Immunopathol* **41**: 153-163.

Martinez D, Vachier N, Stachurski F, Kandassamy Y, Raliniaina M, Aprelon R & Gueye A (2004) Nested PCR for detection and genotyping of *Ehrlichia ruminantium*: use in genetic diversity analysis. *Annals of the New York Academy of Sciences* **1026**: 106-113.

Mastrorunzio JE, Kurscheid S & Fikrig E (2012) Postgenomic analyses reveal development of infectious *Anaplasma phagocytophilum* during transmission from ticks to mice. *J Bacteriol* **194**: 2238-2247.

Mercado-Lubo R, Gauger EJ, Leatham MP, Conway T & Cohen PS (2008) A *Salmonella enterica* serovar typhimurium succinate dehydrogenase/fumarate reductase double mutant is avirulent and immunogenic in BALB/c mice. *Infect Immun* **76**: 1128-1134.

Musoke AJ, McKeever D & Nene V (1997) Subunit vaccines for the control of tick-borne diseases: implications for the future. *Parassitologia* **39**: 131-137.

Nakanishi N, Abe H, Ogura Y, *et al.* (2006) ppGpp with DksA controls gene expression in the locus of enterocyte effacement (LEE) pathogenicity island of enterohaemorrhagic *Escherichia coli* through activation of two virulence regulatory genes. *Molecular microbiology* **61**: 194-205.

Nelson CM, Herron MJ, Felsheim RF, *et al.* (2008) Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *BMC Genomics* **9**: 364.

Norval RA, Andrew HR & Yunker CE (1990) Infection rates with *Cowdria ruminantium* of nymphs and adults of the bont tick *Amblyomma hebraeum* collected in the field in Zimbabwe. *Vet Parasitol* **36**: 277-283.

Nyika A, Barbet AF, Burrridge MJ & Mahan SM (2002) DNA vaccination with map1 gene followed by protein boost augments protection against challenge with *Cowdria ruminantium*, the agent of heartwater. *Vaccine* **20**: 1215-1225.

Nyika A, Mahan SM, Burrridge MJ, McGuire TC, Rurangirwa F & Barbet AF (1998) A DNA vaccine protects mice against the rickettsial agent *Cowdria ruminantium*. *Parasite Immunol* **20**: 111-119.

Oberem PT & Bezuidenhout JD (1987) The production of heartwater vaccine. *Onderstepoort J Vet Res* **54**: 485-488.

Oberem PT & Bezuidenhout JD (1987) Heartwater in hosts other than domestic ruminants. *Onderstepoort J Vet Res* **54**: 271-275.

Paradkar PN, De Domenico I, Durchfort N, Zohn I, Kaplan J & Ward DM (2008) Iron depletion limits intracellular bacterial growth in macrophages. *Blood* **112**: 866-874.

Pilet H, Vachiere N, Berrich M, *et al.* (2012) A new typing technique for the Rickettsiales *Ehrlichia ruminantium*: multiple-locus variable number tandem repeat analysis. *Journal of microbiological methods* **88**: 205-211.

Postigo M, Taoufik A, Bell-Sakyi L, de Vries E, Morrison WI & Jongejan F (2007) Differential transcription of the major antigenic protein 1 multigene family of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks. *Veterinary microbiology* **122**: 298-305.

Postigo M, Taoufik A, Bell-Sakyi L, Bekker CP, de Vries E, Morrison WI & Jongejan F (2008) Host cell-specific protein expression in vitro in *Ehrlichia ruminantium*. *Veterinary microbiology* **128**: 136-147.

Pretorius A, Liebenberg J, Louw E, Collins NE & Allsopp BA (2010) Studies of a polymorphic *Ehrlichia ruminantium* gene for use as a component of a recombinant vaccine against heartwater. *Vaccine* **28**: 3531-3539.

Prozesky L (1987) Diagnosis of heartwater at post-mortem in ruminants and the confirmation of *Cowdria ruminantium* in mice. *Onderstepoort J Vet Res* **54**: 301-303.

Prozesky L (1987) The pathology of heartwater. III. A review. *Onderstepoort J Vet Res* **54**: 281-286.

Pruneau L, Emboule L, Gely P, *et al.* (2012) Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol Med Microbiol* **64**: 66-73.

Raliniaina M, Meyer DF, Pinarello V, *et al.* (2010) Mining the genetic diversity of *Ehrlichia ruminantium* using map genes family. *Vet Parasitol* **167**: 187-195.

Ramsay G (1998) DNA chips: state-of-the art. *Nat Biotechnol* **16**: 40-44.
Renesto P, Rovey C, Schrenzel J, *et al.* (2008) *Rickettsia conorii* transcriptional response within inoculation eschar. *PLoS One* **3**: e3681.

Rikihisa Y & Lin M (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. *Curr Opin Microbiol* **13**: 59-66.

Rockett JC, Christopher Luft J, Brian Garges J, *et al.* (2001) Development of a 950-gene DNA array for examining gene expression patterns in mouse testis. *Genome Biol* **2**: RESEARCH0014.

Smith LT (1996) Role of osmolytes in adaptation of osmotically stressed and chill-stressed *Listeria monocytogenes* grown in liquid media and on processed meat surfaces. *Appl Environ Microbiol* **62**: 3088-3093.

Totte P, Blankaert D, Marique T, Kirkpatrick C, Van Vooren JP & Werenne J (1993) Bovine and human endothelial cell growth on collagen microspheres and their infection with the rickettsia *Cowdria ruminantium*: prospects for cells and vaccine production. *Rev Elev Med Vet Pays Trop* **46**: 153-156.

Uilenberg G (1982) Experimental transmission of *Cowdria ruminantium* by the Gulf coast tick *Amblyomma maculatum*: danger of introducing heartwater and benign African theileriasis onto the American mainland. *American journal of veterinary research* **43**: 1279-1282.

Uilenberg G (1983) Heartwater (*Cowdria ruminantium* infection): current status. *Adv Vet Sci Comp Med* **27**: 427-480.

Uilenberg G, Camus E & Barre N (1985) [A strain of *Cowdria ruminantium* isolated in Guadeloupe (French West Indies)]. *Rev Elev Med Vet Pays Trop* **38**: 34-42.

Van Amstel SR & Oberem PT (1987) The treatment of heartwater. *Onderstepoort J Vet Res* **54**: 475-479.

Van Heerden H, Steyn HC, Allsopp MT, Zweggarth E, Josemans AI & Allsopp BA (2004) Characterization of the pCS20 region of different *Ehrlichia ruminantium* isolates. *Vet Microbiol* **101**: 279-291.

Van Vliet AH, Jongejan F & van der Zeijst BA (1992) Phylogenetic position of *Cowdria ruminantium* (Rickettsiales) determined by analysis of amplified 16S ribosomal DNA sequences. *Int J Syst Bacteriol* **42**: 494-498.

Van Vliet AH, Jongejan F, van Kleef M & van der Zeijst BA (1994) Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. *Infect Immun* **62**: 1451-1456.

Wakeel A, Zhu B, Yu XJ & McBride JW (2010) New insights into molecular Ehrlichia chaffeensis-host interactions. *Microbes Infect.*

Zweygarth E, Josemans AI, Van Strijp MF, Lopez-Rebollar L, Van Kleef M & Allsopp BA (2005) An attenuated *Ehrlichia ruminantium* (Welgevonden stock) vaccine protects small ruminants against virulent heartwater challenge. *Vaccine* **23**: 1695-1702.